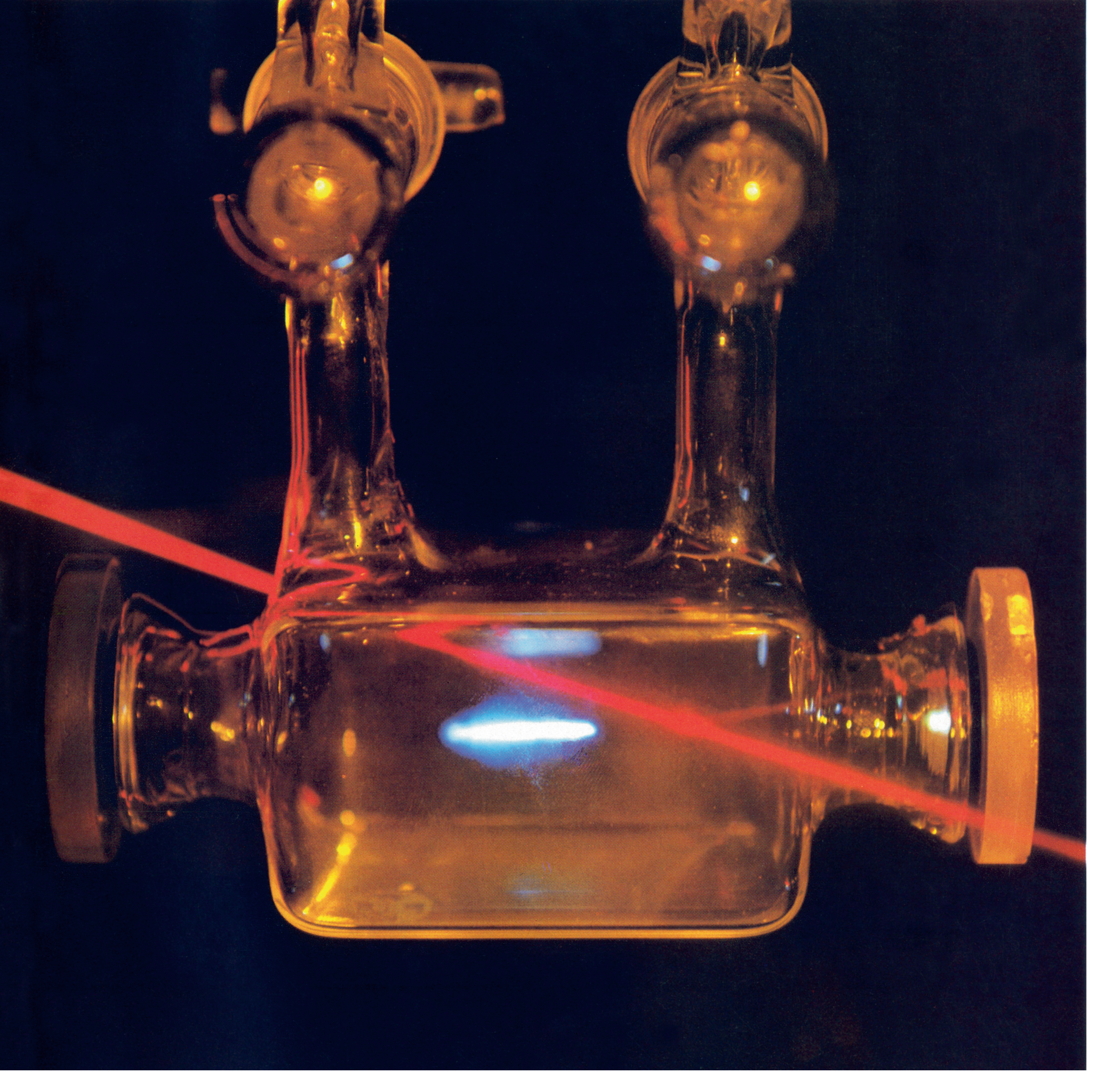
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



LASER EN QUIMICA

Julio 1979

200 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 34
Julio 1979

INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

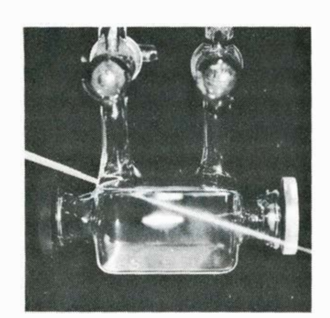
Manuel Puigcerver: *Monóxido de carbono en la galaxia;* Ramón Pascual: *El spin del protón;* Monique Robert: *La base molecular del movimiento celular;* Amando García: *Utilización del láser en química;* Jesús Palomeque: *Un sistema de enfriamiento del cerebro en los mamíferos;* Luis Bou: *Problemas intrínsecamente difíciles;* Julio Samsó: *El cometa de Halley pintado por Giotto;* Luis Bou: *Juegos matemáticos;* José E. Myro y Alicia Menéndez: *Taller y laboratorio.*

Ciencia y sociedad:

Cándido Genovard y Pedro Molera.

Libros:

Rafael Pérez Alvarez-Osorio, José M. Arranz y Mercedes Durfort.



LA PORTADA

La fotografía de la portada muestra la disociación del sulfuro de carbonilo gaseoso (OCS) producida mediante un impulso intenso de un microsegundo de duración procedente de un láser infrarrojo de dióxido de carbono. Las moléculas de dicho compuesto se descomponen por ruptura dieléctrica. Este proceso tiene lugar cuando la intensidad del campo eléctrico asociado con el impulso de un láser alcanza un valor del orden de millones de volt por centímetro. Bajo estas condiciones, las moléculas se rompen [véase "La utilización del láser en química", por Avigdor M. Ronn, en este número]. La luz de color azul-blanco que se observa en el centro de la vasija de reacción está originada por la descomposición violenta del sulfuro de carbonilo que produce el haz focalizado del láser de infrarrojos. (La radiación infrarroja es invisible y no da lugar a ninguna imagen en la fotografía.) La línea de color rojo representa la trayectoria del haz de un láser de helio y neon que actúa como monitor del proceso de descomposición. En el exterior de la vasija de reacción, la trayectoria del haz se hace visible gracias a las partículas de polvo presentes en el aire. En su interior, la trayectoria del haz es invisible hasta que el láser de dióxido de carbono produce la descomposición del sulfuro de carbonilo y el recipiente se llena con partículas de azufre que difunden el haz de helio y neon.

Suscripciones:
Prensa Científica, S.A.
Calabria 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 2200 pesetas
Extranjero:
Un año (12 números): 36 U.S.$
Ejemplar atrasado ordinario:
225 pesetas
Ejemplar atrasado extraordinario:
340 pesetas

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49; Barcelona-10

Distribución para los restantes países:
Editorial Labor, S.A.
Calabria, 235-239; Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovín
Avda. de Moratalaz, 137; Madrid-30
Tel. 430 84 81
Cataluña:
Miguel Munill
Balmes, 191, 2.º, 3.ª; Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Fotografía de la portada de Fritz Goro

Página	Fuente
6-8	Alan D. Iselin
9	Kitt Peak National Observatory (*abajo*), Alan D. Iselin (*arriba*)
10-11	Alan D. Iselin Kitt Peak National Observatory
13	National Radio Astronomy Observatory
14-18	Alan D. Iselin
19	Kitt Peak National Observatory
23	Fritz Goro
24-34	Dan Todd
36-45	Antonio Tovar
53	Elias Lazarides
54-55	Jean Paul Revel
56-57	Ilil Arbel
58	Elias Lazarides (*arriba*), Keigi Fujiwara y Thomas D. Pollard (*abajo*)
60-63	Elias Lazarides
64-66	Ilil Arbel
69	Fritz Goro
70-81	Gabor Kiss
83-84	Carol Donner
85	Mary Ann Baker
86-87	Carol Donner
88	Ilil Arbel
89	Carol Donner
90	Ilil Arbel
93	Patricia J. Wynne
94	Patricia J. Wynne (*arriba*) Albert Miller (*abajo*)
95	Patricia J. Wynne (*arriba*), Albert Miller (*abajo*)
96	Patricia J. Wynne
97-99	Albert Miller
100	Albert Miller
101	Patricia J. Wynne
105	Scala/Editorial Photocolor Archives
106	Scala/Editorial Photocolor Archives
107	Hale Observatories
108	New York Public Library
109	Bettmann Archive (*arriba*); Master and Fellows of Trinity College, Cambridge (*abajo*)
110	National Gallery of Art (*arriba*), Alinari/Editorial Photocolor Archives (*abajo*)
111	Kupferstichkabinett, Berlin (*arriba*); Metropolitan Museum of Art (*abajo*)
112	Dan Todd
115	Ralph Morse
120-125	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición: Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Cayfosa. Santa Perpetua de Moguda
Barcelona
Printed in Spain - Impreso en España

Los autores

M. A. GORDON y W. R. BURTON ("Monóxido de carbono en la galaxia") han colaborado en estudios de estructura galáctica basados en observaciones con radiotelescopios. Gordon es director adjunto del Observatorio Nacional de Radioastronomía (NRAO) y dirige las instalaciones del Observatorio en Tucson y Kitt Peak, Arizona. Estudió Fisica en la Universidad de Yale, y recibió su grado de doctor en astrogeofísica en la Universidad de Colorado en 1966. Sus campos de interés en investigación han incluido las auroras terrestres, radioemisiones de Júpiter, física del medio interestelar y estructura de la galaxia. Burton es profesor de astronomía y director del Departamento de Astronomía en la Universidad de Minnesota. Estudió matemáticas y astronomía en el Swarthmore College y ganó una beca Fulbright para trabajar en astronomía y física en la Universidad de Leiden, donde se doctoró en 1970. Entonces trabajó en las oficinas del Observatorio Nacional de Radioastronomía en Charlottesville, Virginia. Tras ocho años allí, el otoño pasado se trasladó a Minnesota. Sus principales campos de interés son la estructura galáctica y el análisis por ordenador de observaciones radiotelescópicas.

ALAN D. KRISCH ("El spin del protón") es profesor de física de la Universidad de Michigan. Se graduó en la Universidad de Pennsylvania y se doctoró por la de Cornell en 1964. Este mismo año ingresó en la facultad de Michigan. Su investigación se ha centrado en el estudio de las interacciones protón-protón buscando objetos más pequeños dentro del protón. Desde 1972 Krisch se ha concentrado en desarrollar y utilizar el haz de protones polarizados del sincrotón de gradiente cero del Argonne National Laboratory.

ANTONIO TOVAR, ("La conquista de Europa occidental por las lenguas indoeuropeas") nació en Valladolid en 1911. Estudió en las Universidades de Valladolid y Madrid. Fue Becario de la Junta para Ampliación de Estudios. Catedrático de latín en la Universidad de Salamanca, de la que fue Rector. Ha sido profesor de lingüística en las de Buenos Aires y Nacional de Tucumán (Argentina), Universidad de Illinois en Urbana, y actualmente en la Universität Tübingen (Alemania Federal), donde se jubila. Espera colaborar todavía en Madrid en la medida de sus fuerzas en la preparación de nuevas promociones de estudiantes interesados en la lingüística. Ha publicado numerosas obras de filología clásica y de lingüística indoeuropea y de otros campos, como el vasco y lenguas americanas. Es miembro de número de la Real Academia Española y de la de Ciencias de Heidelberg.

ELIAS LAZARIDES Y JEAN PAUL REVEL ("La base molecular del movimiento celular") son biólogos celulares del Instituto de Tecnología de California. Lazarides es profesor adjunto de biología. Griego de nacimiento, se trasladó a los Estados Unidos en 1968 para estudiar química y biología en la Universidad de Wesleyan. Posteriormente, tras graduarse, trabajó en bioquímica en la Universidad de Harvard, doctorándose en 1975. Luego gozó de una beca postdoctoral para trabajar en el laboratorio de Keith R. Porter, de la Universidad de Colorado en Boulder. Se encuentra en el Instituto de Tecnología de California desde 1977. Revel estudió en la Universidad de Estrasburgo. Luego, tras graduarse, se marchó a los Estados Unidos para realizar labor investigadora en bioquímica en Harvard y se doctoró en 1957. Entonces entró a formar parte del departamento de anatomía de Harvard y gradualmente ascendió de colaborador de investigación a catedrático. Se trasladó al Instituto de Tecnología de California en 1971. Los trabajos de Revel se han centrado siempre en la aplicación de la microscopía electrónica al estudio de la ultraestructura de las membranas.

AVIGDOR M. RONN ("Utilización del láser en química") enseña química en el Brooklyn College de la Universidad de Nueva York. Nació en Tel Aviv, donde completó su educación superior. Desde 1956 hasta 1958 sirvió en el cuerpo de artillería de las Fuerzas de Defensa de Israel. A continuación se desplazó a los Estados Unidos con el fin de estudiar química en la Universidad de California en Berkeley. En el año 1966 obtuvo su doctorado en la Universidad de Harvard. Tras disfrutar de una beca postdoctoral en el National Bureau of Standards, se incorporó al Instituto Politécnico de Brooklyn. Ronn ingresó en el Brooklyn College en septiembre de 1973; a partir de esa fecha, su trabajo de investigación se ha llevado a cabo fundamentalmente en relación con los láseres de infrarrojos y la separación de isótopos mediante el láser.

MARY ANN BAKER ("Un sistema de enfriamiento del cerebro en mamíferos") enseña ciencias biomédicas y biología en la Universidad de California en Riverside. Se licenció en fisiología en la Universidad de California en Santa Barbara y se doctoró en anatomía por la Universidad de California en Los Angeles, en 1968. Después de estar dos años con una beca postdoctoral en la Universidad de Washington entró a formar parte del profesorado de fisiología de la Facultad de Medicina de la Universidad de California Meridional. Se trasladó a Riverside en 1976 como uno de los primeros miembros del profesorado del Programa de Ciencias Biomédicas.

LARRY J. STOCKMEYER Y ASHOK K. CHANDRA ("Problemas intrínsecamente difíciles") son especialistas en computación del Thomas J. Watson Research Center, de la International Business Machines Corporation. Stockmeyer es miembro de plantilla del departamento de ciencias matemáticas. Realizó su primer ciclo universitario en matemáticas en el Instituto de Tecnología de Massachusetts, centrando su interés por la teoría de computabilidad en ese mismo centro, una vez graduado. Se recibió de doctor de ingeniería eléctrica en 1974. Sus principales campos de investigación son la teoría de complejidad de computación y los estudios teóricos de almacenamiento de datos. Stockmeyer es también instrumentista de trompeta en un conjunto de jazz formado por ingenieros y científicos de IBM. Chandra es director del proyecto de Teoría de Programación, también en el Watson Research Center. Nacido en la India, obtuvo su título de bachiller en ingeniería eléctrica en el Instituto Indio de Tecnología de Kanpur. Trasladado a los Estados Unidos, recibió el grado Master por la Universidad de California en Berkeley y su doctorado, por la Universidad de Stanford, en 1973.

ROBERTA J. M. OLSON ("El cometa de Halley pintado por Giotto") es profesora adjunta de Historia del Arte en Wheaton College y está especializada en arte renacentista italiano. Estudió en Carleton College, en la Universidad Hebrea de Jerusalén y en la Universidad de Oxford. Obtuvo el título de graduada en 1969 en St. Olaf College con un Phi Beta Kappa. Se licenció en Historia del Arte en la Universidad de Iowa y obtuvo su M.F.A. y su doctorado en la misma disciplina en la Universidad de Princeton (este último título en 1976). La tesis doctoral de Olson versó sobre la última etapa en la pintura de Sandro Botticelli.

Hace...

José María López Piñero

...cuatrocientos años

Falleció el humanista aragonés Juan Lorenzo Palmireno, que durante un cuarto de siglo había ocupado cátedras de poesía, griego, oratoria, retórica y gramática en la Universidad de Valencia.

Palmireno se interesó por las ciencias principalmente a través de su relación con los profesores de medicina y matemáticas de la citada Universidad. La influencia más temprana fue la del médico y naturalista Pedro Jaime Esteve, en cuyos trabajos botánicos llegó a colaborar. Aparte de referencias dispersas en su abundante producción escrita, Palmireno dedicó dos libros a las cuestiones relacionadas con la ciencia. Ambos responden a una intención de vulgarizar los conocimientos científicos y técnicos, poco frecuente en la literatura didáctica de la época. El primero es una *Sylva* (1563), destinada a enseñar a los "niños de gramática" las nociones elementales sobre pesos, medidas y monedas. Mayor interés tiene el segundo, titulado *Vocabulario del humanista* (1569), que intenta ofrecer al "estudioso" una síntesis atractiva de la historia natural y otros saberes afines. Incluye nueve partes, de las cuales tres están consagradas a los animales, dos a las plantas y drogas, otras dos a los metales, piedras preciosas y monedas, y uno a "los términos de escribir". Con la excepción de una, son meros "abecedarios"; la exposición por orden alfabético se complementa con "descansos" sobre un tema concreto, para hacer más amena la lectura. Lo mismo que los libros del sevillano Pedro Mexía y otras obras de divulgación científica de este siglo, el *Vocabulario* de Palmireno depende en buena parte todavía, por su contenido y su formato, de las enciclopedias bajomedievales.

...trescientos años

Joan d'Alós Serradora fue nombrado titular de la Cátedra de Prima de Medicina de la Universidad de Barcelona. Sólidamente instalado en los ambientes oficiales –llegó también a protomédico del Principado de Cataluña y de los Condados de Rosellón y de Cerdeña–, no fue uno de los *novatores* españoles que rompieron durante estos años con la ciencia tradicional, sino un ecléctico ampliamente abierto a las novedades. Publicó tres obras consagradas a temas farmacológicos, pero debe su relieve histórico al libro titulado *De corde hominis disquisitio physiologico-anatomica* (1694), la primera monografía moderna consagrada en España al aparato cardiovascular.

De corde expone, en primer término, la anatomía del corazón, teniendo en cuenta los hallazgos más recientes y desmintiendo las opiniones de los antiguos acerca de la naturaleza de las fibras cardíacas y la unión de los vasos al corazón fetal. La parte fisiológica del libro no es una mera defensa de la doctrina de la circulación de la sangre, ya que incorpora las complejas cuestiones planteadas durante los tres cuartos de siglo que habían transcurrido desde la publicación de la obra de William Harvey. Por ello, discute extensamente cuestiones como el origen de la sangre, las relaciones entre respiración y circulación, y la naturaleza de la contracción cardíaca; adopta, en general, una actitud de compromiso entre las interpretaciones tradicionales y las nuevas teorías.

Muy notable es la parte terapéutica del tratado, que incluye excelente exposición de dos novedades terapéuticas de la época: la inyección endovenosa y la transfusión sanguínea. Detalla las técnicas de administración intravenosa de diferentes sustancias en los animales y en el hombre, anotando las experiencias de varios autores, en especial las del alemán Johann Daniel Major. También explica muy detenidamente las técnicas de transfusión sanguínea de Richard Lower y de otros médicos británicos e italianos, contraponiendo esta nueva posibilidad con el antiguo uso supersticioso de la sangre.

Todavía más importante es la sección consagrada a la anatomía patológica cardiovascular. Recoge cerca de un centenar de protocolos de autopsia referentes a lesiones del corazón y de los vasos, en su mayor parte aparecidos de obras del siglo XVII. Alós añade, además, dos procedentes de su propia experiencia disectiva.

...doscientos años

Juan Justo García terminó de redactar sus *Elementos de Aritmética y Algebra*. Tras ser juzgada favorablemente por una comisión, la obra fue impresa en 1782 y se utilizó como libro de texto en varias universidades de la Península y de la América española. Alcanzó cinco ediciones, apareciendo la última de ellas en 1821-22. La primera es un volumen de casi quinientas páginas, que se inicia con un "Resumen histórico del origen, progreso y estado actual de las matemáticas puras"; incluye, después, una síntesis de aritmética y unos "Principios de Algebra" de mayor extensión, donde se estudia el cálculo lineal y las ecuaciones de primero y segundo grado, a continuación, se ocupa de la geometría gráfica euclídea, la trigonometría plana y, tras unos "Elementos de Geometría práctica" o topografía, de la trigonometría esférica; la última parte ofrece los principios de geometría analítica, cálculo diferencial y cálculo integral. Cuesta Dutari afirma que, a pesar de sus limitaciones, "hay que considerar una hazaña el intento de conducir, en un libro de esta extensión, a los lectores españoles de 1782 desde las matemáticas de la escuela primaria hasta las ecuaciones diferenciales y la que, por esos años, se llamó en Europa la "geometría sublime", es decir, la "geometría diferencial".

Juan Justo García era extremeño y se formó en la Universidad de Salamanca, de la que fue catedrático de matemáticas a partir de 1774. Durante el reinado de Fernando VII, a causa de sus ideas liberales, los absolutistas lo depuraron por vez primera en 1816. El trienio liberal le permitió publicar varias obras, entre las que destaca *Elementos de verdadera lógica* (1821), donde resumió la obra de Destutt-Tracy. Fue entonces diputado por Extremadura en las Cortes y vicerrector de la Universidad de Salamanca. Al volverse a implantar el régimen absolutista, fue de nuevo "impurificado".

...cien años

Falleció en Barcelona el farmacéutico Joaquim Balcells Pascual, una de las figuras que mantuvieron el cultivo de las ciencias experimentales en la Cataluña de mediados del siglo XIX. Se formó en los Colegios de Farmacia de San Victoriano, de Barcelona, y de San Fernando, de Madrid, así como en las escuelas de la Junta de Comercio de Barcelona. Fue catedrático de física, primero, en la Universidad de Cervera, luego, en las

escuelas de la citada Junta de Comercio y, por último, en la Escuela de Ingenieros Industriales de Barcelona. Aunque se ocupó de temas de física en varios trabajos –entre ellos, en una *Memoria sobre las varias clases de vapor* (1833)–, la mayor parte de su producción estuvo dedicada a la química. Publicó estudios aplicados a la agricultura, la industria vinícola y la farmacia, y otros de carácter básico como una *Lhytologia meteorica* (1854) consagrada a la descripción y análisis químico de los aerolitos, con motivo de los que habían caído en varias localidades de Tarragona.

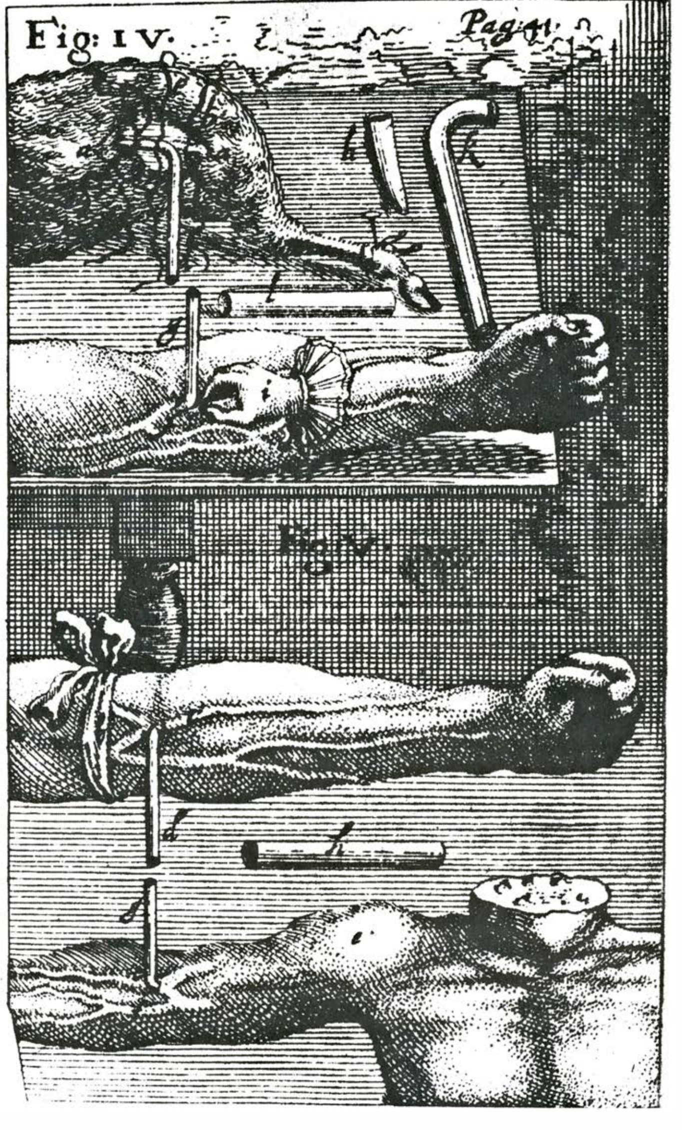
Familiarizado con las técnicas de observación microscópica, Balcells fue uno de los primeros autores europeos que publicaron descripciones precisas del vibrión colérico. Con motivo de la epidemia que asoló a Europa entre 1854 y 1856, realizó observaciones que expuso en las revistas *El Restaurador Farmacéutico* y *El Porvenir Médico*. Describió con notable precisión "vibriones dotados de una movilidad asombrosa", y de "movimiento anguloso como el rayo". En la misma epidemia llevaron a cabo observaciones similares el italiano Filippo Paccini y los portugueses Costa Simoes y Macedo Pinto. Sin embargo, ninguna de ellas tuvo la transcendencia científica que tres décadas más tarde alcanzó la de Robert Koch.

...cincuenta años

Pedro de Novo y Fernández Chicarro fue nombrado profesor de geología en la Escuela de Ingenieros de Minas de Madrid. Tras graduarse en este mismo centro, había trabajado en la Comisión del Mapa Geológico de España junto a Lucas Mallada. Realizó entonces investigaciones en Alicante y Andalucía que expuso, entre otras publicaciones, en el trabajo *De Sierra Morena a Sierra Nevada (Reconocimiento orogénico de la región bética)* (1926). También fue enviado por la citada Comisión a los Estados Unidos y a México, donde estudió la industria y los yacimientos petrolíferos; a su regreso, dirigió los primeros trabajos sistemáticos de localización y ejecución de perforaciones exploratorias que en esta especialidad se realizaron en nuestra Península. Junto a Joaquín de Mendizábal, dirigió, en 1934, una expedición a Guinea Ecuatorial, fruto de la cual fue el primer mapa geológico del territorio.

Novo contribuyó notablemente a la difusión de las modernas ideas geológicas en España con una traducción ejemplar, que apareció en 1923, del monumental tratado *Der Anlitz der Erde,* del austríaco Eduard Suess. Muy interesado por la terminología científica, colaboró en el *Diccionario tecnológico* de Leonardo Torres Quevedo y publicó un *Diccionario de voces usadas en geografía física* (1949).





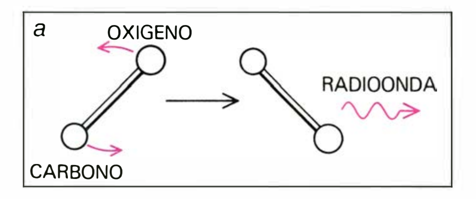
Grabados de la época de Alós relativos a las técnicas de inyección endovenosa y transfusión sanguínea que expone en su libro **De corde hominis**

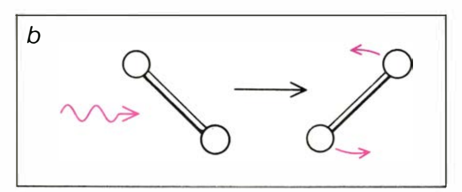
Monóxido de carbono en la galaxia

Los radiotelescopios sintonizados a la longitud de onda característica de emisión de las moléculas de monóxido de carbono revelan la presencia de un anillo de nubes frías generadoras de estrellas en el invisible interior de la Vía Láctea

M. A. Gordon y W. B. Burton

La exploración de nuestra galaxia, la Vía Láctea, es una empresa comparativamente reciente en astronomía. El problema residía en que los árboles ocultaban el bosque, es decir, en la incapacidad de percibir visualmente la estructura en gran escala de la galaxia a través de dos de sus principales componentes: las nubes de gas y el polvo. El advenimiento de la radioastronomía, hace unos cuarenta años, proporcionó un camino para orillar el problema. Las radioondas, que son emitidas por los átomos y moléculas de gas en toda la galaxia, penetran las nubes interestelares mucho más fácilmente que las ondas luminosas, permitiendo a los radioastrónomos cartografiar los caracteres estructurales invisibles de la galaxia. Hasta ahora, la principal longitud de onda usada para esta finalidad ha sido la señal de 21 centímetros emitida por el hidrógeno atómico (H). Presentaremos aquí los resultados de la más moderna fase de esta exploración, que está en gran parte basada en las ondas de longitud mucho más corta emitidas por las moléculas interestelares de monóxido de carbono (CO). Resulta que la representación de la galaxia deducida de las moléculas de monóxido de carbono es muy diferente de la deducida de los átomos de hidrógeno.





UNA MOLECULA de monóxido de carbono emite una radioonda de longitud 2,6 milímetros cuando la molécula cesa de girar a la menor velocidad permitida por la mecánica cuántica (*a*). La molécula de monóxido de carbono puede absorber radiación a la misma longitud de onda iniciando de nuevo su giro a la misma velocidad mínima (*b*).

Ha pasado escasamente más de media centuria desde que los astrónomos llegaron al acuerdo general de admitir que la Vía Láctea es sólo uno de los casi incontables sistemas estelares, o galaxias, algunos de los cuales contienen hasta 300.000 millones de estrellas. La aceptación de este hecho fue difícil, costando casi doscientos años de observación y discusión. Aunque los mayores telescopios del siglo XIX eran capaces de resolver manchas dispersas de nebulosidad en estrellas aisladas, los tamaños y distancias de esas asociaciones estelares eran objeto de debate. Algunos astrónomos mantenían que nuestra galaxia era la única existente, constituyendo el universo entero, y que las asociaciones estelares que se veían alrededor no eran más que pequeñas y relativamente próximas zonas de activa formación de estrellas. Otros estaban de acuerdo con el punto de vista originalmente sugerido en 1755 por Manuel Kant: algunas de las nebulosas son "universos aislados" enormemente distantes, comparables en tamaño y estructura con la Vía Láctea.

La resolución experimental de la disputa se produjo poco después del descubrimiento de que ciertas estrellas variables regularmente pulsantes, las Cefeidas, podían servir como indicadores fidedignos de distancia astronómica. En 1925, Edwin P. Hubble anunció la observación de estrellas cefeidas en Messier 31, la Gran Nebulosa de Andrómeda. La relación característica entre la luminosidad absoluta y el período de pulsación de esas estrellas variables mostró de manera irrefutable que tales nebulosas están en realidad extremadamente lejanas, y de aquí que deban ser consideradas como galaxias por derecho propio.

Una vez que se aceptó la existencia de los universos aislados de Kant, se puso en marcha el estudio de las galaxias individuales, incluyendo la nuestra. Se observó que las galaxias que incorporan cantidades importantes de gas y polvo además de estrellas, tienen frecuentemente una estructura espiral. Hubble ideó un esquema de clasificación en el que las galaxias espirales están dispuestas partiendo de las que muestran una estructura espiral menos desarrollada (designadas S0), pasando por una secuencia de fases intermedias (Sa, Sb, Sc) hasta aquellas con una estructura más desarrollada (Sd). Aunque las galaxias S0 se llaman algunas veces "tempranas" y las Sd "tardías", nada se sabe en realidad acerca de la relación, si es que hay alguna, entre la prominencia de la estructura espiral de una galaxia y su edad. (Para complicar la tarea de clasificación, aproximadamente un tercio de las galaxias espirales presentan una estructura en forma de eje a través de su centro. Estas galaxias se llaman espirales barradas y se designan SB0, SBa, etcétera.)

Se prodría pensar que la galaxia de la Vía Láctea fuese un lugar ideal para iniciar el estudio de las galaxias espirales, ya que el sistema solar está inmerso en ella. No lo es, por la misma razón por la que ha sido tan difícil de estudiar ópticamente: el disco de cualquier galaxia espiral contiene una gran cantidad de gas y polvo (hasta un 10 por ciento de la masa). En un disco visto de canto, el gas y el polvo atenúan eficazmente la luz de las estrellas de la galaxia. El hecho de que el sistema solar queda dentro de tal disco queda de manifiesto cuando se representa la distribución de estrellas visibles dentro de la galaxia.

Aunque la presencia de todo este gas y polvo en un disco plano indica que la galaxia de la Vía Láctea es probablemente del tipo espiral (por referencia a la situación en las galaxias externas), la atenuación resultante limita la investigación óptica convencional de la galaxia a unos 6000 años-luz. La galaxia de la Vía Láctea, sin embargo, parece tener un diámetro de más de 100.000 años-luz. Ni

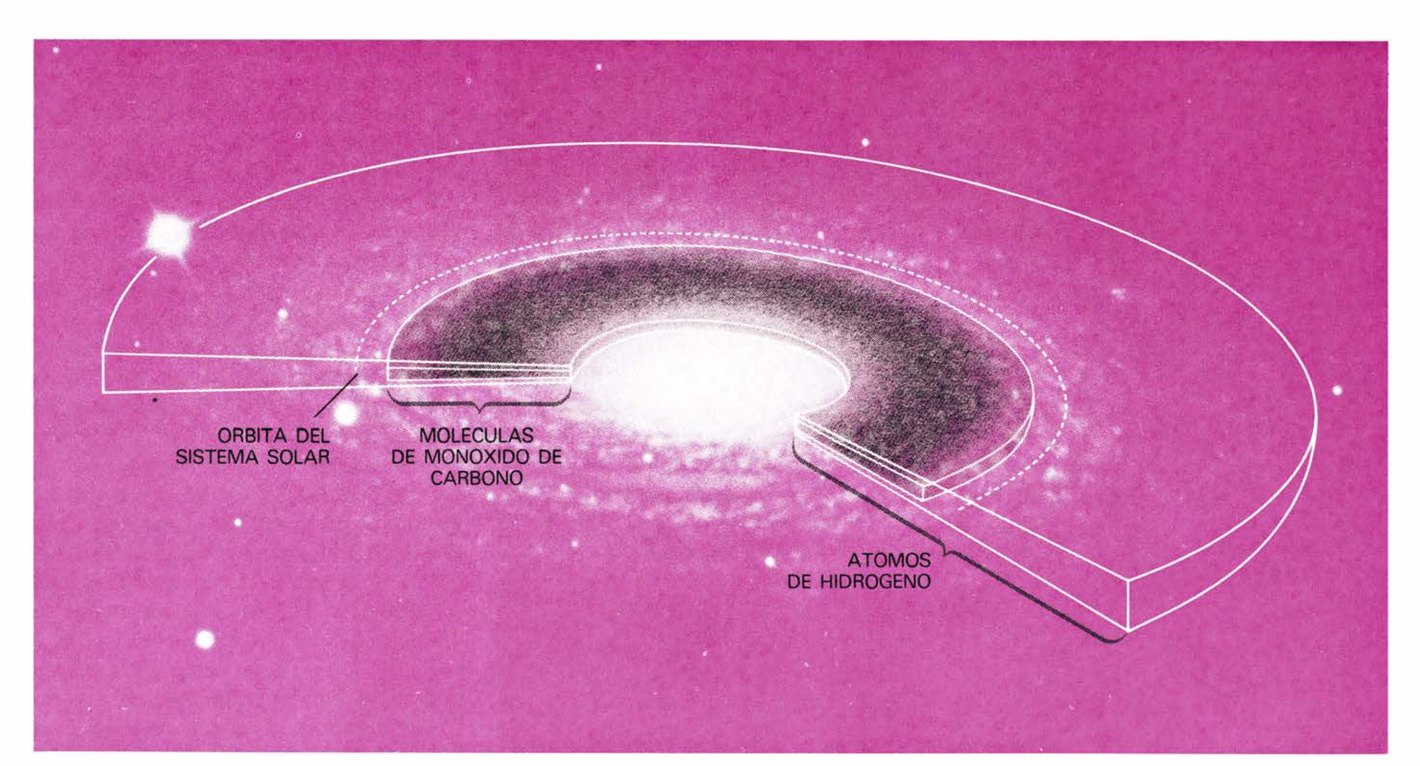
siquiera se puede determinar por medios ópticos la clase de galaxia espiral que es la Vía Láctea, y mucho menos determinar los detalles de su estructura general y composición.

La nueva herramienta que ha capacitado a los astrónomos para trabajar con las galaxias apareció en escena en 1951. En un mismo número de la revista *Nature* en ese año, tres grupos, uno americano, otro holandés y un tercero australiano anunciaron la detección de radioondas procedentes de átomos de hidrógeno interestelar a una longitud de onda de 21 centímetros (correspondiente a una frecuencia de 1420 megahertz). Las radioondas de esta longitud tienen dos propiedades características: su propagación a través del espacio queda comparativamente poco afectada por el gas y polvo interestelar, y las partículas que emiten la mayor parte de la radiación a esta frecuencia, es decir, los átomos de hidrógeno, se sabe están entre los más corrientes componentes del gas interestelar. Por tanto, estudiando la emisión en 21 centímetros, los radioastrónomos pueden determinar la distribución del hidrógeno atómico componente del gas interestelar a grandes distancias, mucho más allá de la región accesible a los telescopios ópticos. Hasta hace poco tiempo, era corriente la creencia de que el hidrógeno atómico de la galaxia revelaría la distribución de los otros componentes del gas interestelar, pero esta hipótesis ha resultado ser errónea.

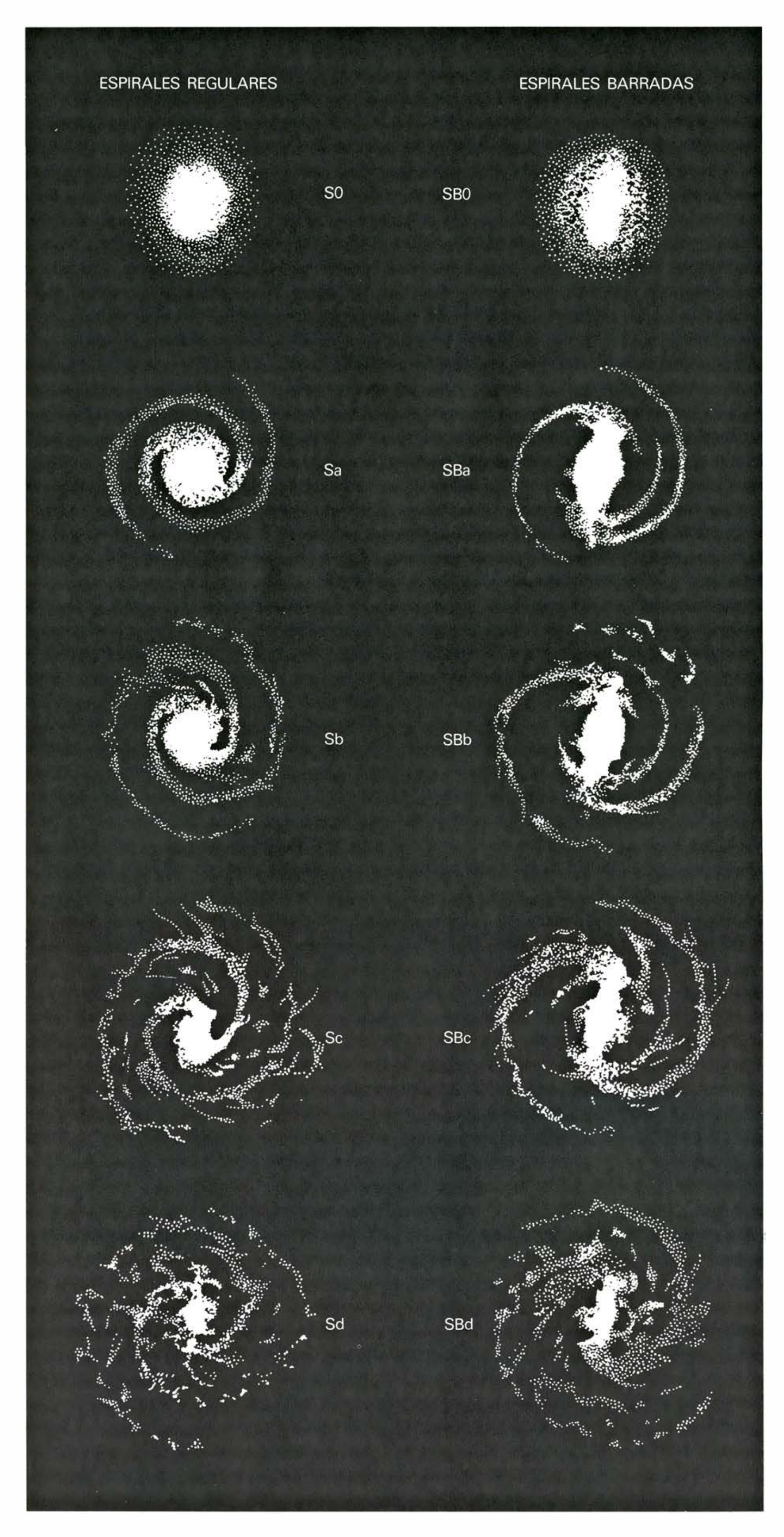
En general, las radioondas procedentes de una clase particular de átomo, se emiten a frecuencias precisas, determinadas por las características del átomo. Como cada uno de los átomos del gas está en movimiento, sin embargo, un observador detecta la onda a una frecuencia ligeramente diferente, corrida en un sentido u otro respecto a la frecuencia característica según sea el movimiento relativo del átomo respecto al observador. Esta propiedad, llamada efecto Doppler, significa que los átomos que se están alejando del observador parecen emitir a frecuencias más bajas, mientras que los que se mueven hacia él parece que emitan a frecuencias más altas. Comparando la frecuencia observada con la característica de la emisión, medida en el laboratorio, el astrónomo puede calcular la velocidad de los átomos de gas respecto al telescopio. La velocidad que aleja el átomo del telescopio se llama positiva y la que lo acerca negativa.

Para explorar la galaxia, el radioastrónomo empieza por medir primero la intensidad de radiación de un átomo en función de la frecuencia. La representación resultante de intensidad en función de la frecuencia constituye un radioespectrograma, y el proceso de efectuar esta clase de observación es la radioespectroscopía. En la práctica, el astrónomo sintoniza el radiorreceptor a la frecuencia característica. Las señales de radio recibidas se llevan a un dispositivo que mide simultáneamente la intensidad a muchas frecuencias diferentes. Para simplificar la interpretación, un ordenador unido al receptor corrige automáticamente los espectrogramas respecto a aquellos corrimientos de frecuencia debidos a velocidades conocidas, tales como la rotación del telescopio alrededor del centro de la Tierra, el movimiento orbital de la Tierra alrededor del Sol e incluso el movimiento del Sol respecto a las estrellas próximas. El ordenador rotula la escala de frecuencia del espectrograma en función de la velocidad media de las estrellas en la vecindad del Sol, sistema de referencia inercial que los astrónomos llaman el sistema local en reposo.

En astronomía, las distancias son generalmente difíciles de medir. Un posible indicador de la distancia es la velocidad del gas en observación. Para cualquier visual desde el telescopio a través del plano de la galaxia de la Vía Láctea, sería deseable poder relacionar la velocidad a la distancia desde el sistema solar. De esta manera, se podría convertir la intensidad de un espectrograma a una velocidad dada en la cantidad de gas existente a la correspondiente distancia a



ANILLO DE MOLECULAS alrededor del núcleo de la galaxia de la Vía Láctea, que se extiende de un radio galáctico de 12.000 años-luz a uno de 24.000 años-luz, alcanzando su máxima densidad a aproximadamente 17.000 años-luz (*zona gris*). En contraste, el radio de la órbita del Sol alrededor del centro galáctico es de 30.000 años-luz (*línea de trazos*). La capa de nubes frías generadoras de estrellas definida por la observación de las moléculas de monóxido de carbono parece tener un espesor de unos 300 años-luz. El anillo molecular queda dentro de un disco de gas interestelar mucho mayor y más espeso, compuesto principalmente de átomos de hidrógeno, que radia a una longitud de onda característica de 21 centímetros. Se estima que el radio total de la galaxia es de unos 50.000 años-luz. El fondo coloreado de la figura es una fotografía de una galaxia espiral comparable.



LA CLASIFICACION DE LAS GALAXIAS sigue un esquema propuesto originalmente por Edwin P. Hubble en 1926. Los diferentes tipos de galaxias que aparecen en este dibujo están ordenados de acuerdo con la prominencia de su estructura espiral. De la información recogida por los radioastrónomos a lo largo de los últimos 25 años poco más o menos, la Vía Láctea es o bien una galaxia Sb o bien una Sc. Hoy se discute si la región nuclear de la galaxia está barrada, como en algunas otras, o no.

lo largo de la visual. A diferencia del problema de convertir radiofrecuencia en velocidad, sin embargo, el presente problema no es inmediato; requiere encontrar previamente una relación adecuada entre velocidad y distancia.

Pero ocurre que la propia galaxia suministra una manera de relacionar la velocidad del gas interestelar con la distancia al sistema solar, haciendo posible explorar las regiones interiores de la galaxia por medio de radioespectrogramas. La situación es como sigue. Supongamos que la galaxia girase como un disco sólido. Cada punto situado sobre una visual a través del disco se movería entonces exactamente a la misma velocidad respecto al telescopio. No habría manera de relacionar velocidad con distancia.

Pero en realidad, las galaxias giran diferencialmente. En otras palabras, la velocidad angular de rotación del gas interestelar y del polvo varía con la distancia al centro galáctico, decreciendo hacia fuera de manera no uniforme. Consecuentemente, la velocidad del gas a lo largo de la visual también cambia con la distancia al sistema solar. En particular, la velocidad del gas a lo largo de la visual aumenta con la distancia hasta el punto en que el radio de la galaxia es perpendicular a la visual y decrece después. En la intersección de la visual con la órbita del Sol en el lado de la galaxia más alejado, la velocidad se hace cero respecto a la velocidad del Sol. Alejándose todavía más del Sol, la velocidad se hace crecientemente negativa respecto a la del Sol. La rotación diferencial de la galaxia es la clave que nos permite relacionar la velocidad del gas representado por una característica espectral dada con la distancia al Sol.

En realidad, la cosa no es tan sencilla. Para cada visual a través de la parte interna de la galaxia, hay generalmente dos posiciones que se mueven a la misma velocidad respecto al Sol: una en la parte próxima del punto perpendicular y otra en la parte lejana. Ambas posiciones, sin embargo, están a la misma distancia del centro de la galaxia. Convirtiendo simplemente velocidad en distancia al centro galáctico más bien que en distancia al Sol se puede evitar la ambigüedad en la distancia.

Basándose en esta técnica, los grupos que trabajaban en Holanda y en Australia combinaron sus datos en 1965 para producir un mapa de la distribución del hidrógeno atómico en la galaxia [*véase la ilustración de la página 10*]. Para construir el mapa, tenían que determinar distancias a lo largo de la visual. La ambigüedad en distancia se resolvió



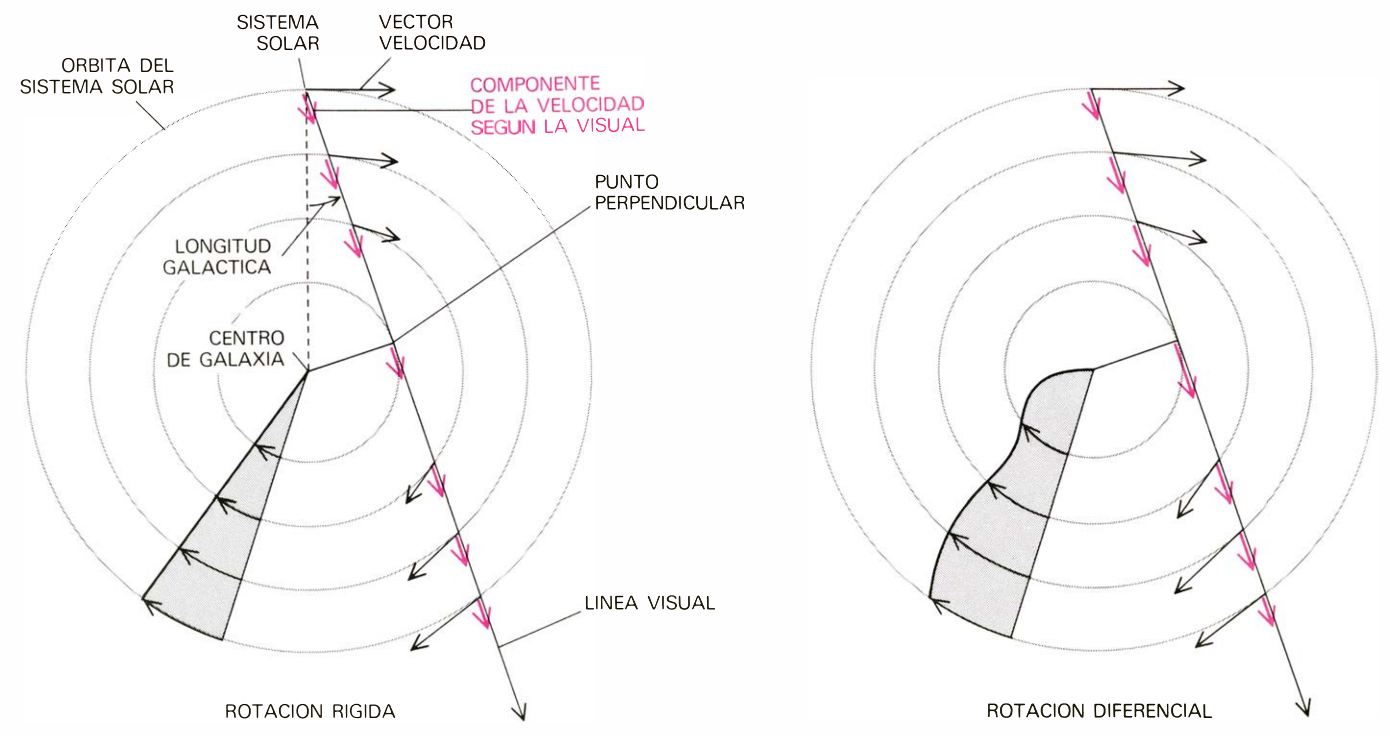
VISTAS "DE CANTO", las galaxias espirales tales como ésta de disco tipo Sb, NGC 4565 en la constelación Coma de Berenice, se distinguen por la eficaz extinción de la luz de las estrellas por nubes de gas y polvo interestelar. La fotografía se hizo con el telescopio de 2,1 metros de Kitt Peak. La distribución de las estrellas en la galaxia de la Vía Láctea muestra que el sistema solar está inmerso en un disco similar de gas y polvo.

considerando muchos espectrogramas dispuestos en láminas perpendiculares al aplastado disco galáctico y tratando de seguir la pista de una característica espectroscópica dada de un espectrograma a otro. Una característica que se extendía sobre muchas visuales se suponía estar hacia el lado próximo del punto perpendicular; una que se extendiera solamente a unas pocas visuales adyacentes se suponía estar en el lado lejano. El mapa resultante no reveló con definida claridad la estructura de la galaxia. O bien nuestra galaxia es apreciablemente más irregular de lo que se suponía, o el método de resolver ambigüedades en la distancia no era efectivo. A pesar de ello, el mapa constituyó uno de los primeros cuadros de la estructura general de la galaxia de la Vía Láctea.

En la proximidad del sistema solar, las investigaciones radioastronómicas de la estructura galáctica no han eliminado en manera alguna los estudios ópticos. La piedra angular del método radioastronómico es el cambio de velocidad con la distancia, cambio que es extremadamente pequeño dentro de los 5000 años-luz del Sol. En esta vecindad, las técnicas ópticas son ampliamente superiores porque no dependen de la velocidad relativa. Recientemente, W. E. Herbst, de la Wesleyan University, examinó todos los "trazadores" ópticos, objetos que incluyen estrellas jóvenes superluminosas, asociaciones de estrellas en regiones polvorientas y nébulas gaseosas.

Su mapa muestra "calles" de estrellas en la vecindad del Sol. Desgraciadamente, la eficaz extinción de las ondas luminosas por el polvo interestelar limita el tamaño de su mapa y por tanto todo conocimiento independiente de la estructura general de la galaxia.

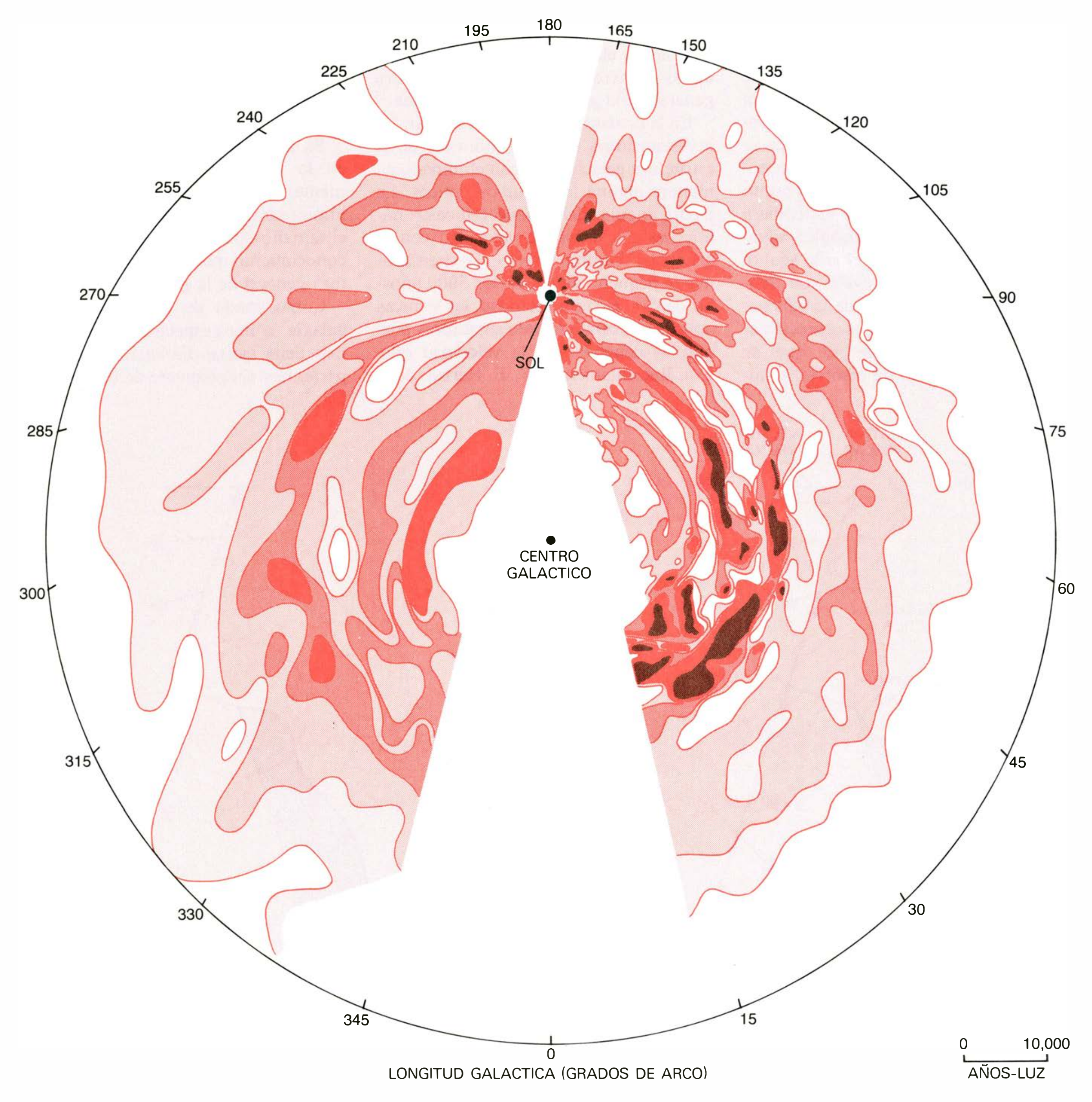
Como sonda de la estructura de la galaxia, la línea espectral de 21 centímetros tiene ciertas limitaciones. Por una parte, hay simplemente demasido hidró-



ROTACION DIFERENCIAL de la galaxia. Constituye un factor clave en la exploración de la estructura galáctica por radioondas. En las dos situaciones representadas aquí, la circunferencia más externa marca la órbita del sistema solar alrededor del centro de la galaxia y el área sombreada representa el movimiento de la materia interestelar dentro de esta órbita durante cierta fracción del año galáctico. Si la galaxia girase como un disco rígido, como se supone en la figura de la izquierda, cualquier punto a lo largo de una visual a través del disco se movería exactamente con la misma velocidad respecto al radiotelescopio, y no habría manera de relacionar la velocidad con la distancia al sistema solar. El hecho de que la galaxia gira más rápidamente en las regiones interiores, como se ve en el diagrama de la derecha, significa que la velocidad del gas a lo largo de la visual cambia en función de la distancia al telescopio. Para la visual particular elegida en este caso, la velocidad de rotación del gas aumenta con la distancia hasta el punto en que el radio galáctico es perpendicular a la visual, y disminuye a partir de allí. En la intersección de la visual con la órbita del sistema solar en el borde externo de la galaxia, esta velocidad se hace nula respecto a la velocidad del sistema solar; más allá, la velocidad se hace crecientemente negativa. El vector componente de la velocidad a lo largo de la visual (flechas de color) se determina midiendo el corrimiento de la longitud de la onda característica con respecto a su valor obtenido en el laboratorio (resultado del efecto Doppler). La longitud galáctica se mide en grados de arco.

geno atómico en el gas interestelar para proporcionar un cuadro neto. Por otra parte, la diminuta masa de los átomos de hidrógeno los hace particularmente vulnerables a fluctuaciones producidas por efectos térmicos. Esta agitación ensancha la frecuencia de radioemisión, enturbiando con ello el movimiento en gran escala de la galaxia que es la base de la estimación de distancias. Además, en las regiones de formación de estrellas que son de las más interesantes de una galaxia, la mayor parte del hidrógeno se encuentra en forma de moléculas diatómicas (H_2), y el hidrógeno molecular no se puede detectar mediante radiotelescopios. Aunque las radioondas de los átomos de hidrógeno son adecuadas para penetrar el polvoriento horizonte de la galaxia, la estructura galáctica deducida de observaciones de hidrógeno atómico sólo es insatisfactoria.

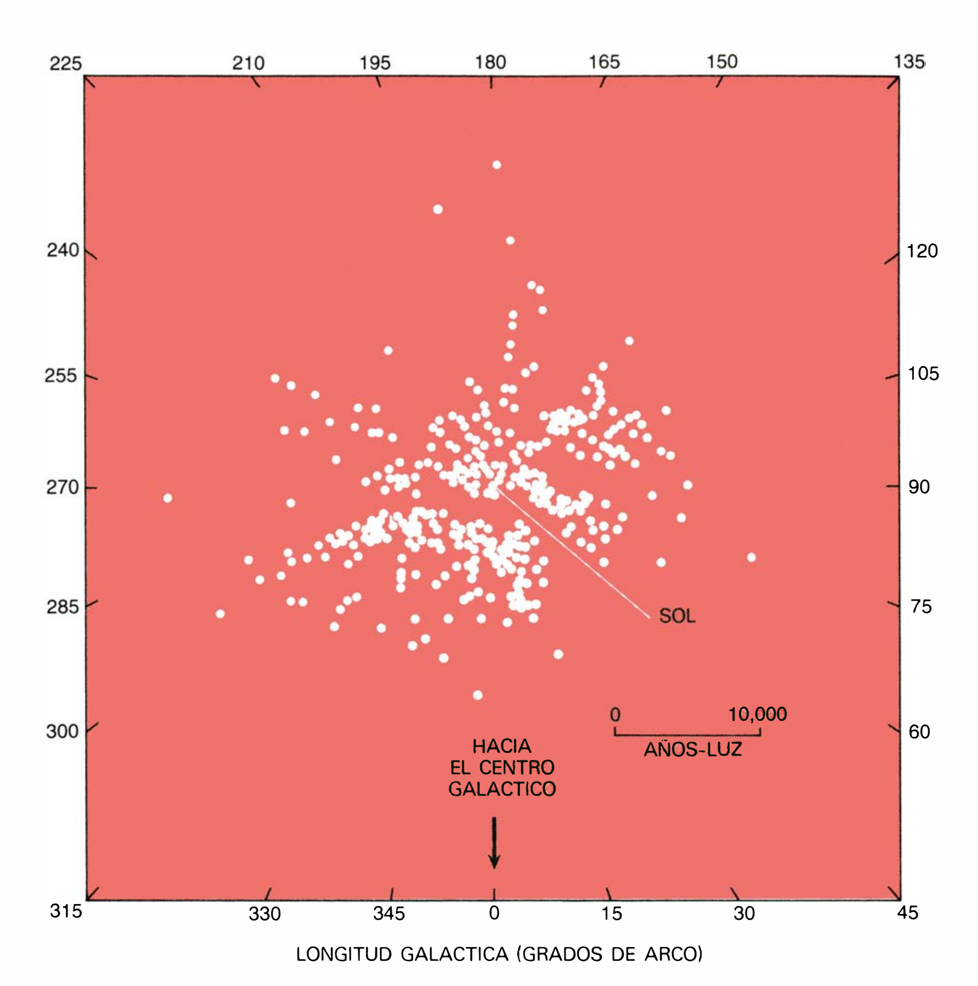
El reciente descubrimiento de las radioondas procedentes del monóxido de carbono interestelar suministra la clave para una más amplia comprensión de la estructura de la galaxia de la Vía Láctea. Las fotografías de galaxias espirales tomadas en planta muestran que las características más conspicuas de estas galaxias son las estrellas extremadamente jóvenes, las nebulosas emisoras y las calles de polvo, todas las cuales delinean la estructura espiral de las galaxias. Esas características coinciden con regiones de formación de estrellas, donde las nubes de gases se hacen frías y densas y los vértices turbulentos en su interior se reúnen para formar estrellas. Las regiones de formación de estrellas son también típicamente asiento de



PRIMITIVA REPRESENTACION de la estructura de la galaxia de la Vía Láctea basada en un gran número de observaciones radiotelescópicas hechas a fines del decenio de 1950 a partir de las radioondas de 21 centímetros emitidas por los átomos de hidrógeno distribuidos a lo largo de una amplia diversidad de longitudes galácticas. La representación combina observaciones hechas en Holanda (la mayor parte por Gart Westerhout y Maarten Schmidt) con las realizadas en Australia (la mayoría por Frank J. Kerr). Las cuñas vacías en el dibujo corresponden a regiones donde la técnica analítica no se puede aplicar porque la componente según la visual de la velocidad atribuible a la rotación de la galaxia es cercana a cero.

moléculas y no átomos de hidrógeno. De aquí que la detección de intensas y ubicuas señales de radio procedentes de una de esas moléculas, el monóxido de carbono, sea un descubrimiento excitante. Significa que las regiones de formación de estrellas se pueden ahora investigar directamente a escala galáctica.

La molécula de monóxido de carbono emite su señal fundamental de radio a la muy corta longitud de onda de 2,6 milímetros. El proceso de emisión es el siguiente: La molécula de monóxido de carbono se puede considerar como un resonador que consta de un átomo de carbono y un átomo de oxígeno conectados eléctricamente. (En realidad, sólo el núcleo de los átomos forma el resonador; los electrones asociados a los átomos rodean el resonador como un enjambre de abejas.) Una molécula de esta clase puede almacenar energía de tres maneras: la vibración de cada átomo respecto al otro, el giro de cada átomo alrededor del otro y los cambios en el tamaño de la nube electrónica. Sin embargo, como la temperatura típica de las nubes oscuras en el medio interestelar es solamente de unos grados sobre el cero absoluto, hay poca energía disponible para almacenar. Por tanto, la principal actividad almacenadora de energía de la molécula de monóxido de carbono en las nubes oscuras lleva consigo el mecanismo de mínima energía, es decir, la rotación de la molécula sobre sí misma.



CALLES DE ESTRELLAS en la proximidad del Sol, evidentes en este mapa en el que la posición de diversos trazadores ópticos se ha representado respecto al plano de la galaxia. El mapa, compilado por W. E. Herbst, incluye nebulosas gaseosas, estrellas jóvenes y brillantes y asociaciones de estrellas en regiones polvorientas. El tamaño de tal mapa queda limitado por la eficaz extinción de la luz de las estrellas por el gas y polvo interestelar. El área del disco galáctico representada en este mapa, por ejemplo, es menos del cinco por ciento del área cubierta por el radiomapa de la página anterior.

La mecánica cuántica da los detalles del proceso. La mínima cantidad de energía que puede almacenarse es igual a la energía que la molécula adquiere cuando pasa de no girar en absoluto a girar a la misma velocidad permitida por la mecánica cuántica, velocidad que está determinada por el momento de inercia de la molécula de monóxido de carbono alrededor de su eje de rotación. Para radiar a una longitud de onda de 2,6 milímetros, la molécula simplemente cesa de girar. Para absorber radiación a la misma longitud de onda, comienza a girar nuevamente. La emisión o absorción de energía puede ser estimulada por choques con otras partículas o por las propias radioondas.

Las ondas de radio procedentes del monóxido de carbono interestelar fueron observadas por primera vez en 1970 por Robert W. Wilson, Keith B. Jefferts y Arno Penzias, de los Laboratorios Bell, trabajando con el radiotelescopio de onda corta de once metros del Observatorio Nacional de Radioastronomía (NRAO) de Kitt Peak, estado de Arizona. Observaciones posteriores revelaron que las señales del monóxido de carbono se podían detectar a lo largo de una amplia extensión de la galaxia. Además, se podían detectar señales procedentes de diferentes isótopos, o especies nucleares, de monóxido de carbono. La relación de intensidades de las señales procedentes de dos de los diferentes isótopos variaba substancialmente, según se encontró, de la relación hallada normalmente en la Tierra. O bien las abundancias isotópicas son muy diferentes fuera del sistema solar, o el monóxido de carbono es tan abundante dentro de las nubes galácticas que sólo una pequeña fracción de su radioenergía puede escapar.

La hipótesis más probable es la última. Si la radiación de las moléculas de monóxido de carbono interacciona estrechamente con los movimientos de la nube gaseosa, su estudio debería suministrar directamente la temperatura de la nube. Por otra parte, como toda la radiación emitida por uno de los isótopos menos abundantes del monóxido de carbono escapa enteramente de la nube, la radiación total es proporcional al número total de emisores dentro de la nube. De aquí que, a diferencia de la radiación del hidrógeno, la procedente de los isótopos del monóxido de carbono se puede analizar para dar tanto la temperatura como la cantidad de gas en regiones distantes e invisibles de la galaxia.

Pasaron más de tres años entre el descubrimiento de las radioondas procedentes del monóxido de carbono interestelar y su introducción como sonda de la estructura galáctica en gran escala. La demora se debió enteramente a dificultades técnicas y no a lentitud en su aceptación por parte de los astrónomos. La frecuencia correspondiente a la longitud de onda de 2,6 milímetros de la señal del monóxido de carbono es de 115.271 megahertz, unas 1000 veces más alta que las frecuencias corrientemente emitidas por las radioestaciones comerciales en frecuencia modulada. Las observaciones a estas frecuencias requieren receptores y antenas complicados. La energía total a medir es pequeña. (Todas las señales no terrestres detectadas durante la totalidad de la historia de la radioastronomía no suman mucho más que la energía liberada por la ceniza de

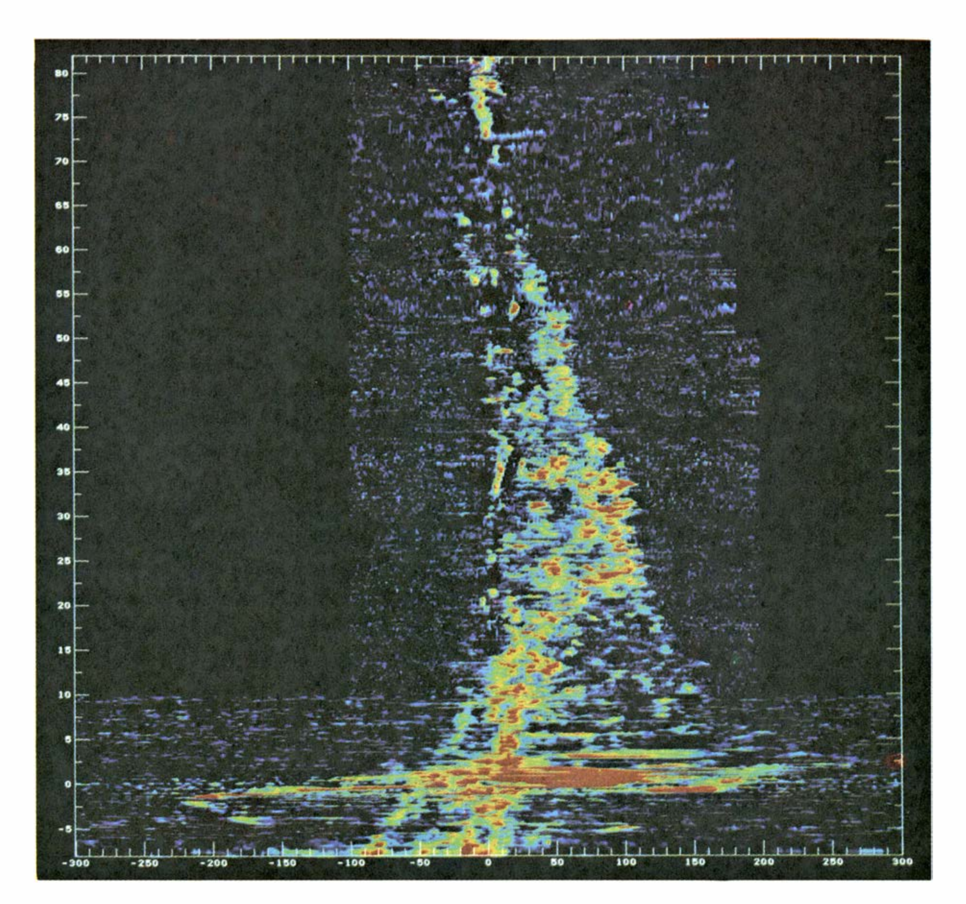
ESTRUCTURA ESPIRAL DE DOS GALAXIAS, claramente delineada por calles de estrellas jóvenes y brillantes y nebulosas gaseosas. Vista más o menos en planta, nuestra galaxia podría parecerse a una de éstas. La fotografía de la parte alta representa una galaxia del tipo Sb, la NGC 3031; la de abajo muestra una del tipo Sc, la NGC 5457. Ambas están en la Osa Mayor. Las fotografías se tomaron con el telescopio de 4 m de Kitt Peak.

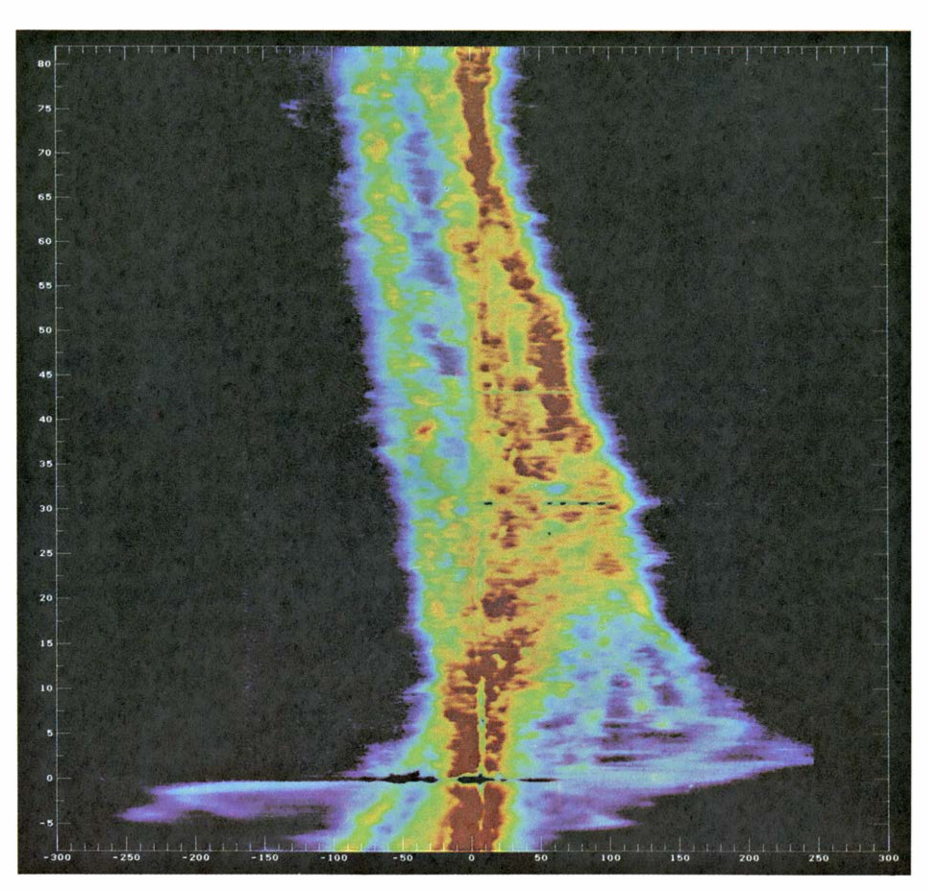
un cigarrillo al caer desde 75 centímetros.)

Las visuales a través del plano central de la galaxia de la Vía Láctea pueden atravesar un dominio de velocidades superiores a 100 kilómetros por segundo, que es equivalente a más de 38 megahertz a la frecuencia del monóxido de carbono. Para separar esta ancha banda de emisión de los efectos instrumentales de fondo, el espectrómetro tendría que cubrir un dominio de velocidades de 300 kilómetros por segundo, que corresponde a un intervalo de frecuencias de 115 megahertz. Además, el radiorreceptor mismo tendría que ser extremadamente sensible y estable. En la época en que se detectó por primera vez la señal del monóxido de carbono, sin embargo, la mayor parte de los radiómetros de ondas milimétricas eran a duras penas adecuados para detectar intensas emisiones de nubes individuales, sin llegar ni con mucho a los requisitos mínimos para la exploración galáctica.

Afortunadamente, los ingenieros del NRAO fueron capaces de satisfacer tales requisitos. El receptor que actualmente se emplea para esta investigación en unión con el telescopio de 11 metros del NRAO en Kitt Peak funciona a 15 grados Kelvin (-258 grados Celsius) para hacer máxima su sensibilidad a las señales radioastronómicas. La temperatura equivalente del ruido térmico generado por este receptor es de unos 200 grados Kelvin, aproximadamente 100 grados por debajo de la temperatura del laboratorio. El receptor y el espectrómetro cubren una anchura de banda de 256 megahertz, que corresponde a un dominio de velocidades de 660 kilómetros por segundo a la frecuencia del monóxido de carbono.

Nuestras observaciones directas del monóxido de carbono en el plano galáctico incluyen espectrogramas registrados cada 0,2 segundos de arco a longitudes galácticas entre 10 y 80 grados, aumentadas por observaciones similares del centro galáctico llevadas a cabo por Thomas Bania, que entonces trabajaba en el NRAO [*véanse las ilustraciones de esta misma página*]. Los espectrogramas de este tipo con frecuencia se alinean verticalmente y se presentan de manera que la intensidad de la señal registrada modula los colores de la ilustración. El diagrama resultante es simplemente una manera conveniente de inspeccionar un gran número de espectrogramas. Aunque contiene la información espectroscópica esencial de la que se deduce la distribución física del gas en la galaxia, no es en manera alguna análogo a una imagen fotográfica de una galaxia externa.





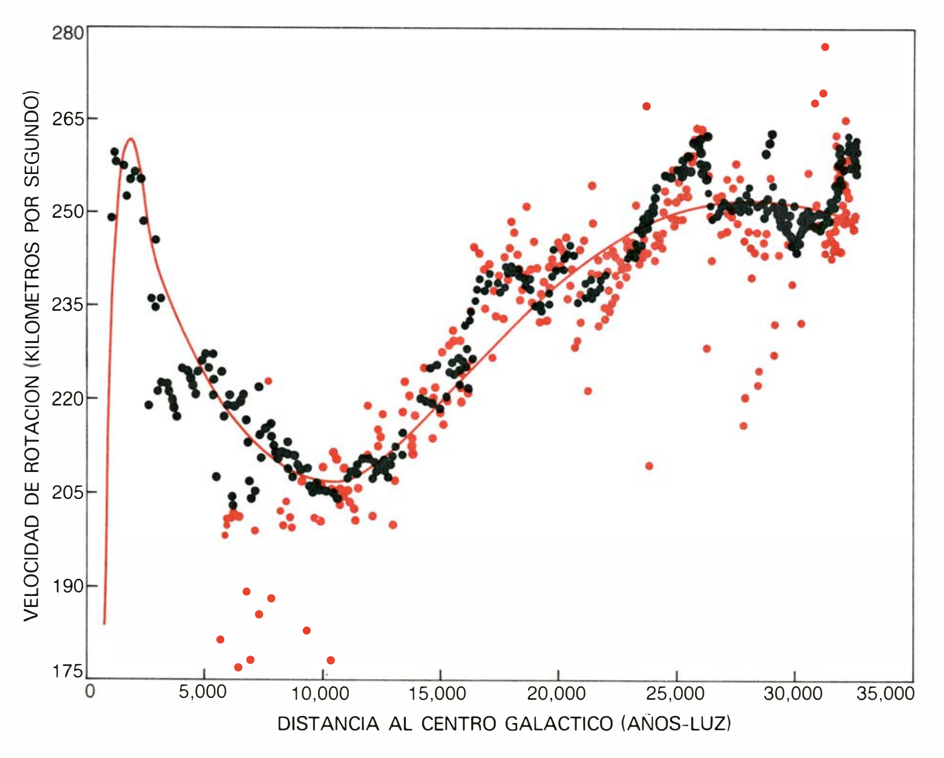
RADIOESPECTROGRAMAS registrados cada 0,2 grados de arco a varias longitudes a lo largo del ecuador galáctico y alineados verticalmente para producir estas dos ilustraciones, en las que la intensidad de la señal registrada modula los colores, variando desde el negro (no hay emisión) hasta el rojo (emisión máxima). Los espectogramas de arriba fueron realizados por Thomas Bania y los autores, basándose en la emisión de 2,6 milímetros de las moléculas de monóxido de carbono; los de abajo fueron hechos por Bania y Westerhout, basándose en la emisión de 21 centímetros de los átomos de hidrógeno. Los datos del CO tienden a estar más agrupados, indicando nubes discretas. La velocidad del gas (*escalas horizontales*) se puede interpretar en términos de distancia desde el centro galáctico.

El diagrama que acompaña a la ilustración representa observaciones comparables de la línea de 21 centímetros del hidrógeno atómico. El mapa se basa en observaciones realizadas por Gart Westerhout con el radiotelescopio de 100 metros del Instituto Max Planck de Radioastronomía en Alemania Occidental y por Bania con el telescopio de 43 metros del NRAO de Green Bank, en Virginia Occidental.

Recordando la geometría de la rotación de la galaxia, es posible aprender mucho acerca de las nubes galácticas sin más que mirar a los datos espectroscópicos no elaborados. Las diferencias más llamativas entre el diagrama del monóxido de carbono y el del hidrógeno son la apariencia extremadamente "picuda" de los datos del monóxido de carbono, la ausencia de monóxido de carbono moviéndose a velocidades negativas a longitudes galácticas mayores de 10 grados y el decrecimiento gradual de la cantidad total de emisión del monóxido de carbono a medida que la longitud aumenta. La apariencia picuda muestra que las radioondas del monóxido de carbono se emiten de manera no uniforme sobre el plano galáctico; presumiblemente, la emisión procede de densas nubes bien definidas en las que pueden estar formándose estrellas.

La geometría de los movimientos observada a lo largo de las visuales muestra que las velocidades negativas que se encuentran a lo largo de visuales que pasan por la parte interna de la galaxia están relacionadas con la materia que queda fuera de la órbita del Sol alrededor del centro de la galaxia. La ausencia de emisión procedente de moléculas de monóxido de carbono que se muevan a velocidades negativas y la disminución de emisión hacia mayores longitudes galácticas indican que el monóxido de carbono está confinado en la región interior de la Vía Láctea. (Esta situación contrasta marcadamente con la del hidrógeno atómico, en el cual la fuerte emisión a velocidades negativas indica abundancia de gas hidrógeno en las regiones externas de la galaxia.) Además, la marcada deficiencia en la emisión de monóxido de carbono a velocidades positivas para longitudes inferiores a 20 grados significa que las moléculas de monóxido de carbono no son abundantes dentro de un radio galáctico de unos 12.000 años-luz. El mismo déficit se observa para átomos de hidrógeno. Evidentemente, tanto las formas atómicas como las moleculares de gas interestelar son escasas en la región central de la galaxia.

Para analizar cuantitativamente las observaciones de monóxido de carbono es necesario convertir la información referente a velocidad en distancia desde el centro galáctico. En otras palabras, es preciso saber cómo gira la galaxia en función del radio galáctico, relación física conocida como curva de rotación galáctica. Como indicamos arriba, la máxima velocidad a lo largo de cualquier visual que pase por el interior de la galaxia es la del material en el punto perpendicular. La distancia entre ese punto y el centro galáctico cambia con la longitud de la visual. El borde derecho de los espectrogramas representa el gas en el punto perpendicular, y la velocidad en ese punto (tras efectuar la corrección debida al movimiento del Sol) da la velocidad de rotación de la galaxia a la correspondiente distancia al centro de aquella. Midiendo esas velocidades positivas máximas en función de la longitud, puede determinarse la curva de rotación.



CURVA DE ROTACION DIFERENCIAL de la galaxia, construida sobre la base de las radioobservaciones resumidas en la ilustración de la página anterior. Las velocidades medidas a lo largo del borde derecho de los espectrogramas en esas ilustraciones se sumaron primero a la componente según la visual de la velocidad orbital del Sol y se representaron entonces en función de la longitud galáctica. Los datos del monóxido de carbono están en color, los del hidrógeno atómico en negro. La naturaleza curvilínea de ambos conjuntos de datos demuestra que la galaxia no gira como un cuerpo rígido.

El precedente análisis está basado en dos hipótesis principales. La primera es que todo el movimiento galáctico es circular: la materia de la galaxia se mueve sólo en órbitas circulares en torno al centro galáctico. En realidad no toda la materia se mueve de esta manera. Hay variaciones en gran escala de la velocidad en torno a la suave curva de rotación que son causadas por "vientos" en el plano galáctico impulsados por la gravedad y por ondas de choque. Al objeto de esta discusión, sin embargo, ninguno de esos dos mecanismos es importante.

La segunda hipótesis es que la materia está tan uniformemente distribuida sobre el plano galáctico que siempre hay gas en los puntos perpendiculares. La distribución de hidrógeno atómico es lo suficientemente ubicua para justificar la hipótesis. Sin embargo, la distribución de monóxido de carbono es discreta. Las regiones vacías entre las bolsas de monóxido de carbono son las causantes de gran parte de la variabilidad en las velocidades medias de las que se deduce la curva de rotación. El examen del diagrama velocidad-longitud para el monóxido de carbono muestra que tal variabilidad no procede de ruido instrumental. La naturaleza de la dispersión suministra información directa sobre la extensión de las regiones vacías y, de aquí, sobre materias tales como la probabilidad de choques entre las bolsas. La dispersión también contiene información sobre los movimientos al azar de las nubes individuales.

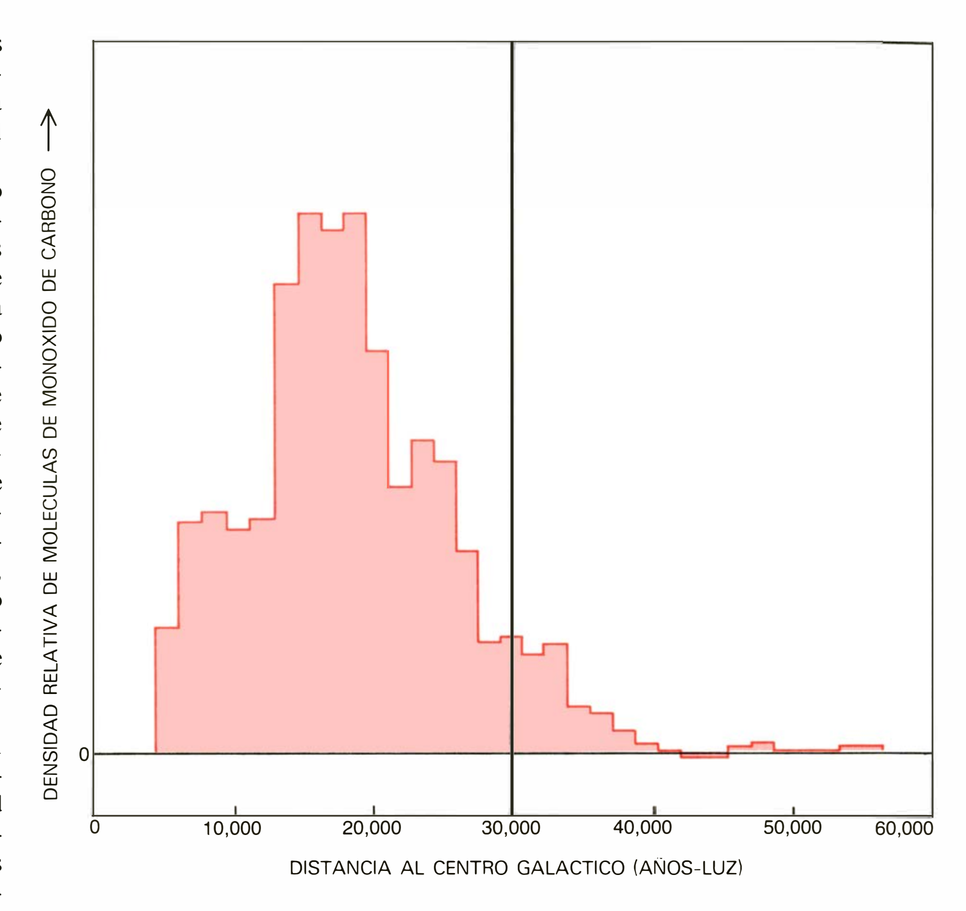
Las moléculas de monóxido de carbono indican regiones de formación de estrellas porque éstas se forman en nubes de gas densas y frías. Los remolinos turbulentos en las nubes pueden producir aumentos locales en la densi-

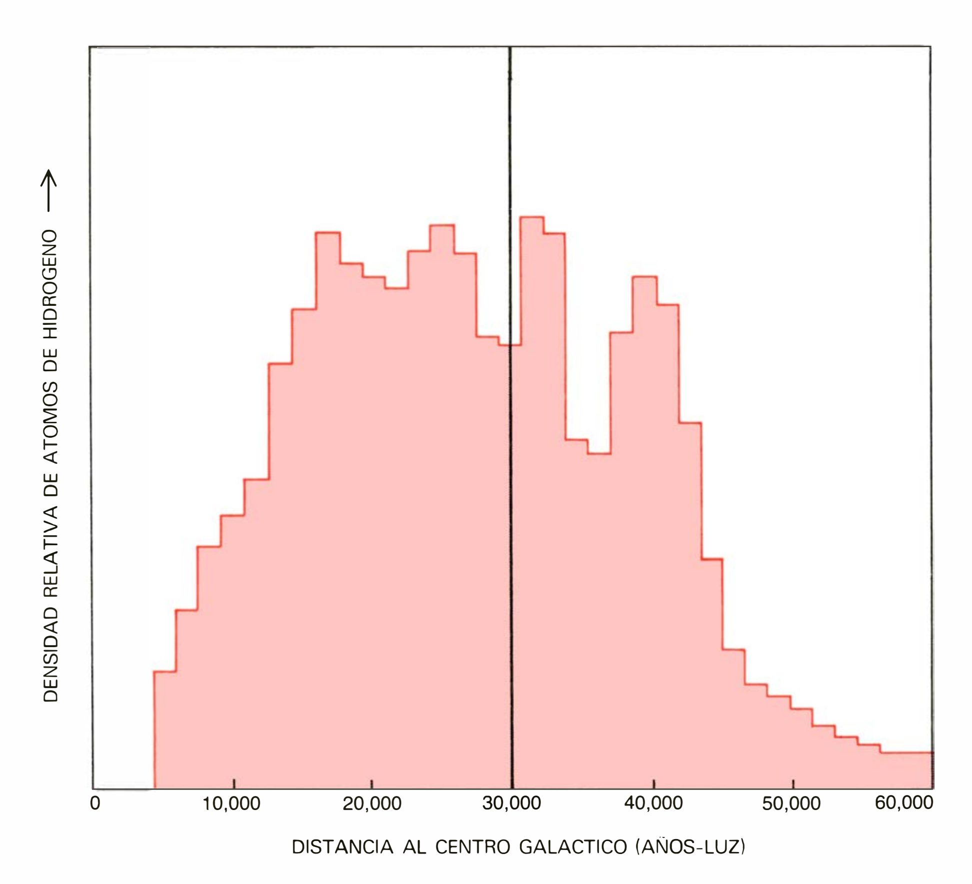
dad del gas hasta el punto en que las moléculas del gas se atraen gravitatoriamente entre sí. Entonces comienzan a condensarse en una protoestrella. Al contraerse el gas también se calienta. Con el tiempo, la temperatura se hace lo suficientemente alta para iniciar un reacción nuclear, que libera la energía de los átomos. Esta energía es lo que hace que una estrella brille y se expansione hasta que la fuerza gravitatoria hacia dentro queda equilibrada por la presión de radiación hacia fuera. Las moléculas de monóxido de carbono están dentro de las densas nubes de gas, que son lo suficientemente opacas para autoprotegerse de la disociación por la radiación ultravioleta procedente de las estrellas calientes recién formadas. Por consiguiente, cartografiar la distribución de monóxido de carbono en la galaxia equivale a cartografiar las regiones de formación de estrellas, activa y potencial, en la galaxia.

El análisis cuantitativo es sencillo. A partir de la curva de rotación de la galaxia, podemos determinar la intensidad total de la emisión del monóxido de carbono (y de aquí el número de nubes frías capaces de formar estrellas) en función del radio galáctico. El procedimiento consiste en considerar intervalos sucesivos a lo largo de una visual dada, calcular la velociad dentro de cada intervalo y preguntar si el espectrograma apropiado muestra las características de emisión de monóxido de carbono a esa velocidad. Si lo hace, la intensidad se resta del espectrograma y se registra en función del radio galáctico del intervalo.

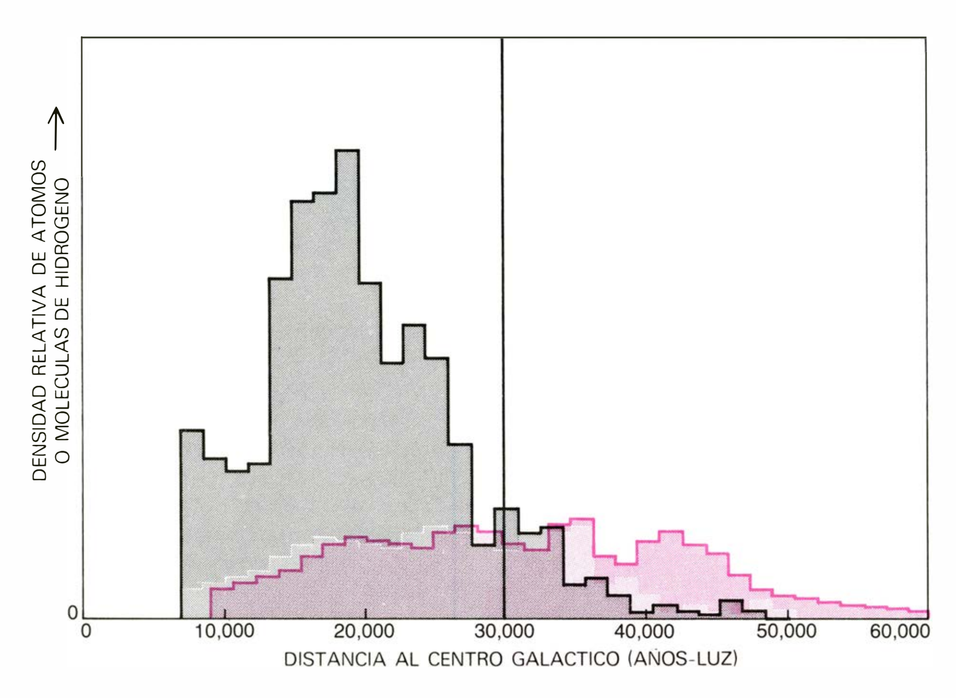
El histograma resultante muestra una sorprendente variación en la intensidad de la emisión del monóxido de carbono con el radio galáctico [*véanse las ilustraciones de esta misma página*]. La mayor parte del monóxido de carbono está dentro de la órbita del Sol, y ésta se halla a 30.000 años-luz del centro galáctico. La cantidad de materia fría aumenta con la distancia hasta un radio galáctico de 12.000 años-luz, donde decrece bruscamente. La máxima intensidad de radiación procedente del monóxido de carbono se encuentra a 17.000 años-luz del centro galáctico. Aceptando que la galaxia es simétrica, el monóxido de carbono se encuentra evidentemente distribuido en un anillo alrededor del centro galáctico con el Sol en la proximidad del borde exterior.

Las moléculas de monóxido de carbono son excitadas por choques con moléculas de hidrógeno; los datos muestran indirectamente, pues, la manera de variar la abundancia de hidrógeno en función del radio galáctico. La relación





DISTRIBUCION de las moléculas de monóxido de carbono (*arriba*) y la de los átomos de hidrógeno (*abajo*) en la galaxia; dicha distribución se representa en estos histogramas en función de la distancia en años-luz al centro galáctico. La línea negra vertical en cada diagrama representa el radio de la órbita del Sol alrededor del centro de la galaxia a la distancia de 30.000 años-luz. La mayor parte del monóxido de carbono está concentrado en un anillo alrededor del centro galáctico con el Sol situado cerca del borde exterior. La intensidad de radioemisión del CO alcanza un máximo a unos 17.000 años-luz del centro galáctico. La distribución de hidrógeno atómico no está confinada a este anillo.



LA DIFERENCIA entre la distribución galáctica del hidrógeno atómico (*color*) y la distribución del hidrógeno molecular (*gris*) es llamativa. Cerca del centro galáctico, la mayor parte del hidrógeno parece estar presente en la forma molecular (H_2) mientras que a distancias mayores que el radio de la órbita solar (*línea negra vertical*) la mayor parte de él parece estar en forma atómica (H). La elaboración de la materia interestelar (formando moléculas complejas, polvo frío y comprimido y, con el tiempo, estrellas) reflejada por esta distribución, tiene lugar en la parte interior del anillo galáctico.

entre el hidrógeno molecular y el monóxido de carbono es íntima, porque el hidrógeno molecular es el único gas presente en la galaxia con densidad suficiente para excitar las transiciones observadas entre los estados de energía rotacional de la molécula de monóxido de carbono. En el gas interestelar, el hidrógeno molecular es aproximadamente 10.000 veces más abundante que el monóxido de carbono, que es el gas siguiente en abundancia.

Uno de los aspectos más importantes de los descubrimientos referentes al monóxido de carbono es lo que nos pueden decir acerca de la cantidad total de hidrógeno molecular en la galaxia. De la absorción conocida de rayos X difusos cerca del núcleo de la galaxia, junto con nuestros datos, se puede estimar la densidad del hidrógeno molecular y compararla con la densidad deducida de un análisis similar referente al hidrógeno atómico. Merced al predominio de hidrógeno en el gas interestelar, la suma de esas dos densidades proporciona una buena estima de la densidad total del gas interestelar. La diferencia de distribución radial entre el hidrógeno atómico y el molecular es sorprendente. A distancias del centro galáctico mayores que la órbita del Sol, el hidrógeno se encuentra principalmente en forma atómica; gradualmente cambia a la forma molecular con la distancia hacia el centro galáctico. Aun cuando la concentración de hidrógeno en cualquiera de sus formas decrece bruscamente dentro de un radio de 12.000 años-luz, el gas allí se encuentra sobre todo en forma de hidrógeno molecular. Como la probabilidad de formación de estrellas es máxima en las regiones frías y densas, ricas en moléculas de hidrógeno, el sistema solar debe quedar en la parte exterior de las regiones de importante formación de estrellas en la galaxia; por tanto, en el disco galáctico, y más allá del centro galáctico de lo que está el Sol, deben formarse pocas estrellas en comparación. (Esta generalización es estrictamente estadística, porque nuestro análisis ha suavizado la distribución de gas en azimut galáctico. No elimina la existencia de manchas aisladas de vigorosa formación de estrellas en las regiones exteriores de la galaxia.) Si un observador se encontrara fuera de la galaxia y mirándola en planta, el Sol aparecería en el borde del disco visible.

Recientemente, varios descubrimientos procedentes de diversas ramas de la astronomía han confirmado esta conclusión. Se suponía antes que la distribución del hidrógeno atómico que revelan los datos de 21 centímetros representaba el prototipo de la distribución de todos los otros componentes del medio interestelar. Ahora está claro que la distribución del hidrógeno atómico es única, definiendo una galaxia mucho mayor que la concebida a partir de los datos moleculares y las regiones asociadas de estrellas muy jóvenes. Se está descubriendo una situación similar en las galaxias espirales externas a medida que la creciente sensibilidad de las medidas de 21 centímetros va haciendo posible cartografiarlas. Los mapas muestran que la región del hidrógeno atómico se extiende hacia fuera al menos hasta dos veces el radio de la imagen fotográfica, que registra la luz emitida por las más brillantes de las estrellas recientemente formadas y por las nubes de gas ionizado que las rodean. La dinámica de esta distribución no se conoce bien. Por ejemplo, no se sabe qué es lo que confina al hidrógeno atómico de las partes exteriores de la galaxia en una delgada capa, manteniéndolo ligado a la galaxia y girando en una órbita circular en torno al centro galáctico.

¿Qué ocurre con el espesor de la capa de monóxido de carbono? Para determinar la cantidad total de gas interestelar hay que medir tanto la variación de átomos de hidrógeno como de moléculas de monóxido de carbono con respecto a la latitud galáctica. Hemos hecho tal medida a lo largo de la longitud galáctica que pasa, muy aproximadamente, a través de la región de máxima intensidad de monóxido de carbono, a 17.000 años-luz del centro galáctico. Los resultados muestran que, aunque el disco de hidrógeno atómico tiene quizá 650 años-luz de espesor, el anillo de monóxido de carbono sólo tiene 300 años-luz, lo que coincide con el espesor estimado de la capa de estrellas muy jóvenes que se ven ópticamente en la vecindad del Sol. Esta es una indicación adicional de la estrecha relación entre las densas nubes de gas y las estrellas jóvenes.

Richard Cohen y Patrick Thaddeus, de la Administración Nacional de Aeronáutica y del Espacio (NASA) y de la Universidad de Columbia, han investigado el espesor de la capa de monóxido de carbono para cierto número de longitudes. Sus descubrimientos muestran que el espesor aumenta desde el centro galáctico hacia fuera hasta el Sol, resultado concordante con el comportamiento medido del hidrógeno atómico, aun cuando la capa de hidrógeno es dos veces más espesa. Como el espesor de la capa de monóxido de carbono suministra información sobre el confinamiento del potencial gravitatorio total al ecuador galáctico, este resultado presenta interés en el contexto de la dinámica galáctica.

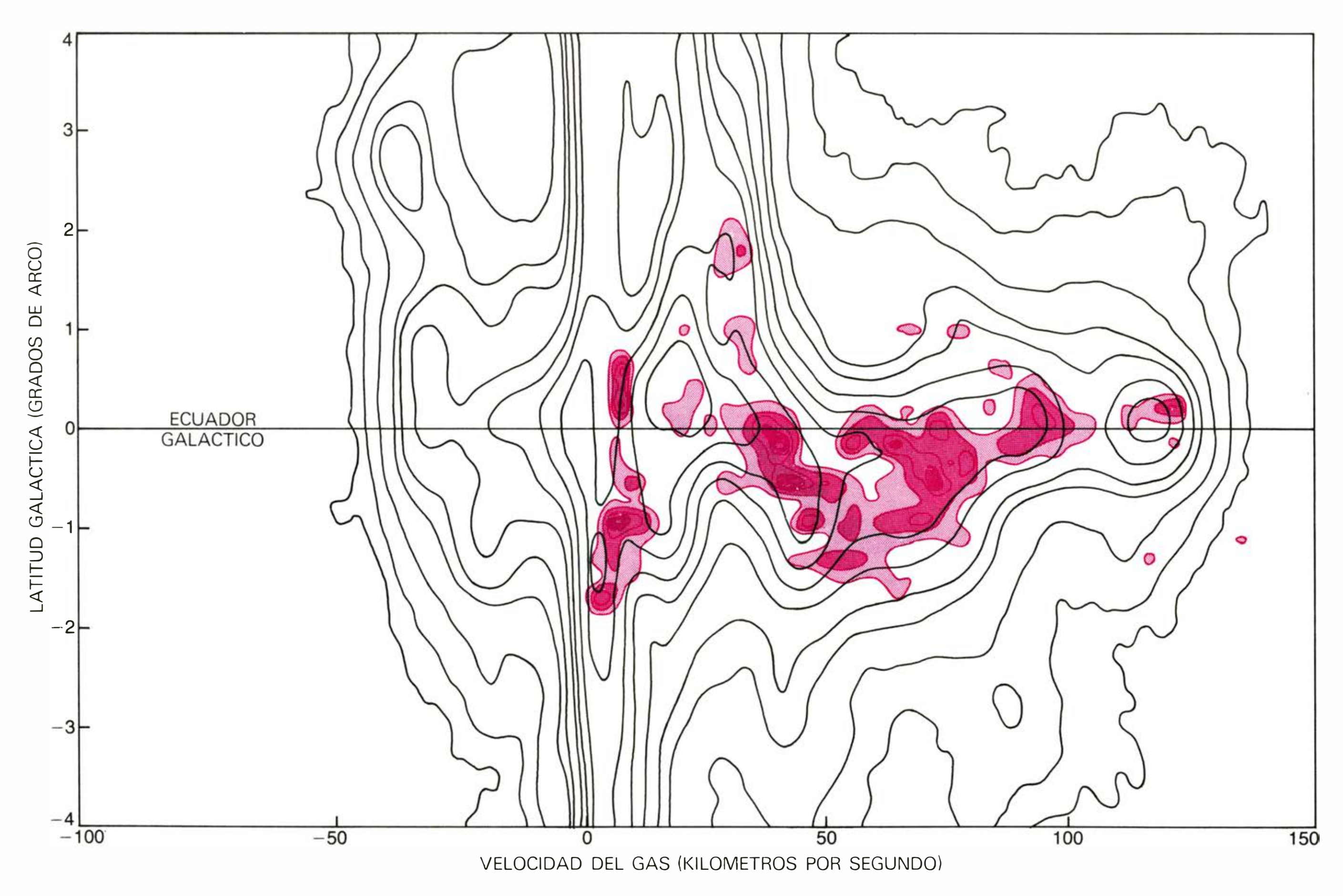
En una galaxia, la relación entre la masa de gas y la masa de estrellas puede revelar mucho acerca de la eficacia del proceso de formación de estrellas. Tomando un valor hipotético para el espesor de la capa de monóxido de carbono y de la capa de hidrógeno, se puede calcular la distribución de la masa gaseosa

(excluyendo las estrellas) en función del radio galáctico, ya que el hidrógeno atómico y el hidrógeno molecular son, con mucho, los componentes gaseosos dominantes en la galaxia. Más difícil resulta calcular la masa total (incluida la de las estrellas). Se han realizado numerosos esfuerzos en ese sentido, combinando observaciones de estrellas en la vecindad del Sol con la rotación galáctica deducida a fin de generar modelos de la distribución de las estrellas invisibles en la galaxia de la Vía Láctea.

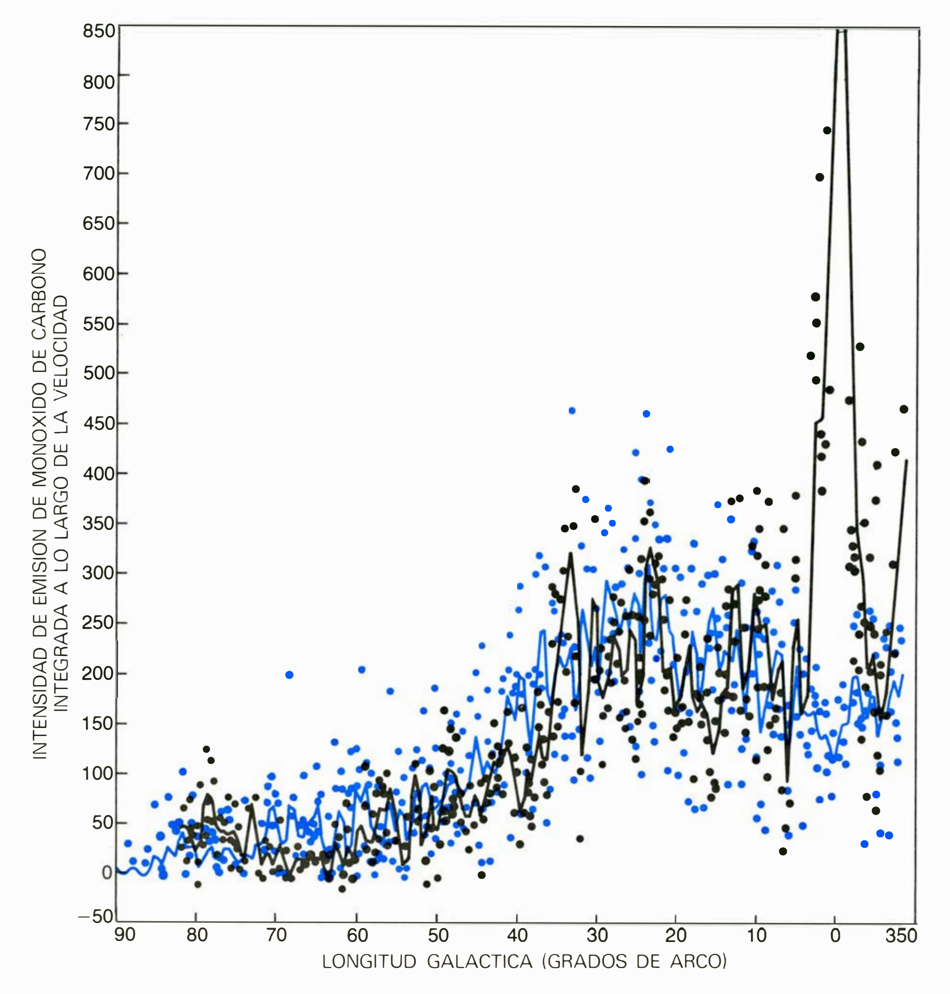
Un reciente modelo teórico presentado por Kimmo A. Innanen, de la Universidad de Nueva York, concuerda sorprendentemente bien con la distribución observada de la masa gaseosa. Los datos de Innanen y nuestros resultados referentes al monóxido de carbono sugieren que a una distancia de entre 30.000 y 12.000 años-luz del núcleo galáctico, la relación de estrellas a gas es aproximadamente de un 4 por ciento en masa, lo que implica una velocidad constante de formación de estrellas en la región intermedia de la galaxia. Este resultado requiere cambios en conceptos anteriores acerca de cómo tiene lugar la formación de estrellas en una galaxia espiral. Mucho de lo que antes se sabía acerca de la distribución de gas en la galaxia procedía de radioobservaciones de hidrógeno atómico, cuya abundancia respecto a las estrellas decrecía con la distancia hacia el centro de la galaxia. La hipótesis subyacente era que la formación de estrellas tenía lugar más vigorosamente en las regiones galácticas centrales, empobreciendo así el hidrógeno interestelar. Los nuevos datos muestran que el gas interestelar no se ha empobrecido; existe en la forma de gas molecular. De hecho, la masa total de hidrógeno molecular en la galaxia es igual a dos billones de veces la masa del Sol y es la misma que la masa total de hidrógeno atómico. Una nueva teoría que explique con éxito la formación de las estrellas debe dar cuenta de esta inesperada constancia de la relación de gas a estrellas.

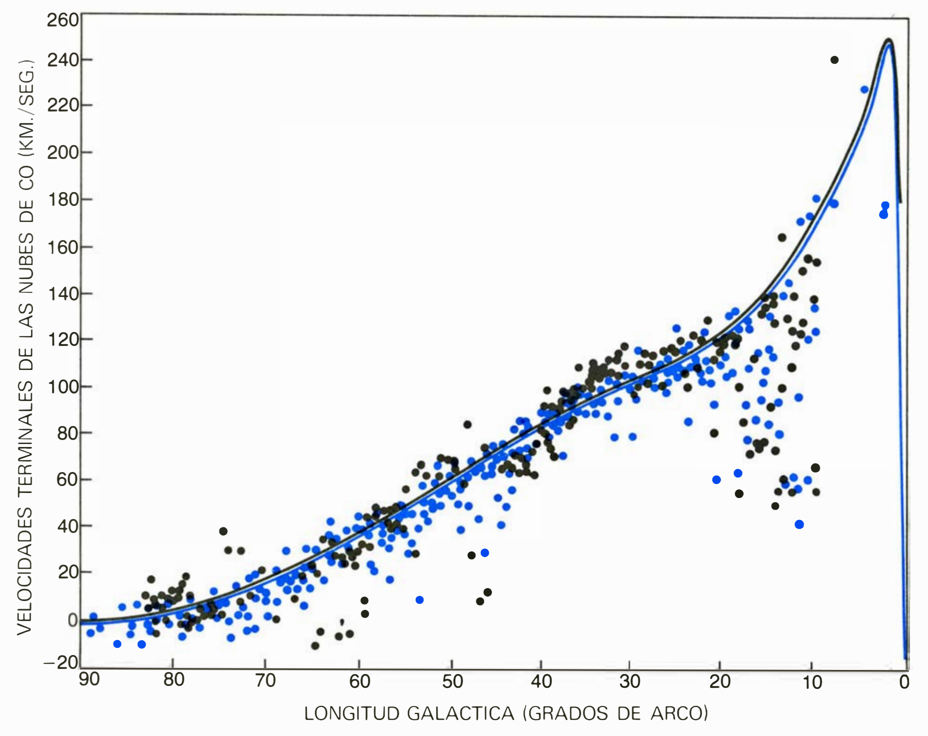
Todo modelo astronómico es una mezcla de hechos observados e hipótesis razonables. En nuestros espectrogramas hemos observado miles de señales características del monóxido de carbono representando un número igual de nubes oscuras. Como solamente podíamos explorar el gas dentro del plano galáctico a intervalos que eran grandes comparados con el tamaño de las nubes, debe haber muchas más nubes oscuras de las que hemos observado, demasiadas para ser consideradas individualmente. Con el telescopio de 11 metros del NRAO haría falta un año para cartografiar completamente la distribución de moléculas en la galaxia de la Vía Láctea. Ello no es posible, entre otras razones porque hay gran demanda para usar el instrumento en otras investigaciones. En lugar de ello, hemos analizado nuestros datos desde un punto de vista estadístico siguiendo lo que se llaman métodos de Monte Carlo. Suministramos a un ordenador variables básicas de las nubes consideradas singularmente, tales como su tamaño, y variables generales de todas las nubes tales como la curva de rotación de la galaxia y el espesor de la capa de monóxido de carbono. También le proporcionamos una primera aproximación de los valores de las variables que deseamos determinar: las velocidades relativas de las nubes cada una respecto a la otra, el número de nubes y su separación espacial.

El ordenador coloca las nubes al azar en el modelo de galaxia, que "observa"



ESPESOR de la capa molecular de la galaxia. En este diagrama se compara con el de la capa de hidrógeno atómico. La ilustración representa la velocidad del monóxido de carbono y del hidrógeno atómico en función de la latitud galáctica a lo largo de visuales adyacentes distribuidas en ángulo recto respecto a la capa gaseosa y pasando muy aproximadamente a través de la región de máxima intensidad de monóxido de carbono a un radio de 17.000 añoz-luz del centro galáctico. Los resultados muestran que la capa de monóxido de carbono (*curvas coloreadas*) tiene aproximadamente la mitad de espesor que la capa de hidrógeno (*negras*). La apariencia más agrupada de los datos de CO sugiere que estas moléculas forman nubes discretas.





AQUI SE HACEN DOS COMPARACIONES entre las radioobservaciones reales de monóxido de carbono en la galaxia (*negro*) y las propiedades predichas por el modelo para ordenador de los autores respecto a la distribución galáctica de las nubes de monóxido de carbono (*color*). El gráfico de arriba, que da la intensidad total de la emisión de monóxido de carbono integrada a todas las velocidades y representada en función de la longitud galáctica, muestra que en este respecto el modelo concuerda bien con las observaciones. El gráfico de abajo, que da las máximas velocidades positivas del monóxido de carbono (es decir, las correspondientes a los puntos perpendiculares) en función de la longitud galáctica, muestra todavía mejor concordancia. La segunda de estas comprobaciones es mucho más sensible que la primera para dar el número de nubes, su separación característica y sus velocidades relativas.

entonces de una manera que imita a nuestras observaciones reales. Comparando éstas con las observaciones del ordenador, podemos someter a comprobación algunas de nuestras hipótesis. Alterando ligeramente las variables y las hipótesis, podemos ver cuál es la sensibilidad del modelo a nuestras hipótesis iniciales y, consiguientemente, averiguar hasta qué punto nuestras estimas de los valores eran buenas. Una alta sensibilidad favorece un buen resultado y viceversa. El manejo de varias variables conduce no sólo a nuevos horizontes respecto a las características en gran escala de las nubes oscuras en la galaxia, sino también a nueva información acerca de la estructura de ésta.

El éxito final de nuestro modelo se puede juzgar examinando la ilustración de esta misma página. La familia de curvas de arriba presenta la comparación más sencilla entre las observaciones y el modelo: compara las intensidades de las emisiones observadas y simuladas de monóxido de carbono, integradas sobre todas las velocidades y representadas en función de la longitud. Las dos curvas concuerdan bien, demostrando que hemos modelado correctamente las nubes de monóxido de carbono a lo largo de cada visual no sólo en cuanto a número, sino también en cuanto al grado de aleatoriedad (como muestra el margen de dispersión).

Otra manera de probar el modelo es examinar las máximas velocidades positivas en función de la longitud galáctica. Este ensayo es mucho más sensible que el descrito arriba porque somete a prueba no sólo las características de velocidad sino la distribución de nubes en el modelo. El hecho de que las velocidades medias encajen en las observaciones indica que la curva de rotación supuesta es la correcta.

A longitudes galácticas entre 25 y 10 grados, la dispersión en la velocidad positiva máxima aumenta bruscamente. En esas direcciones, la velocidad radial varía rápidamente en la vecindad del punto perpendicular. Como el número de nubes oscuras decrece en la parte interior del anillo de monóxido de carbono y como la velocidad radial cambia rápidamente allí, la probabilidad de que una nube se encuentre en el punto perpendicular se hace pequeña. Las nubes alejadas del punto perpendicular generalmente tendrán velocidades más bajas y sus máximas velocidades caerán por debajo del valor idealizado que se muestra. Una situación similar se aplica a las longitudes superiores a 55 grados. Aquí las nubes se hacen tan escasas que la probabilidad de que una

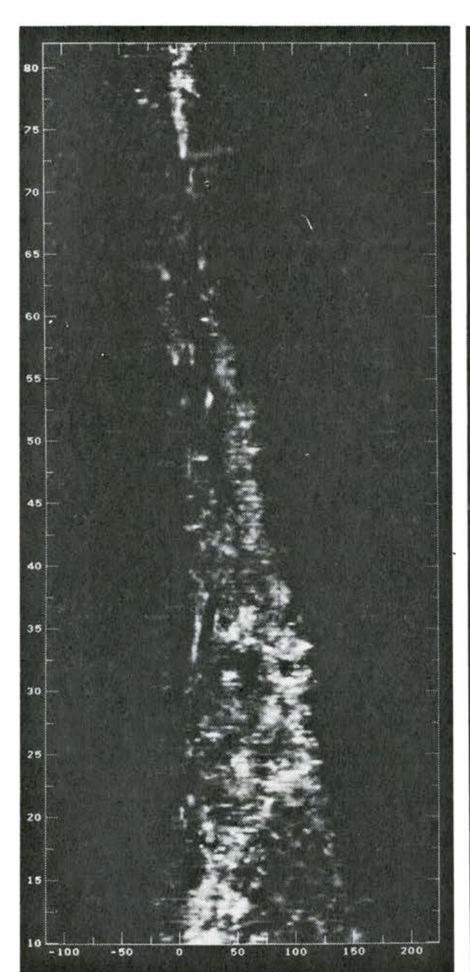
de ellas esté en el punto perpendicular es pequeña. En ambas regiones el modelo concuerda bien con las observaciones.

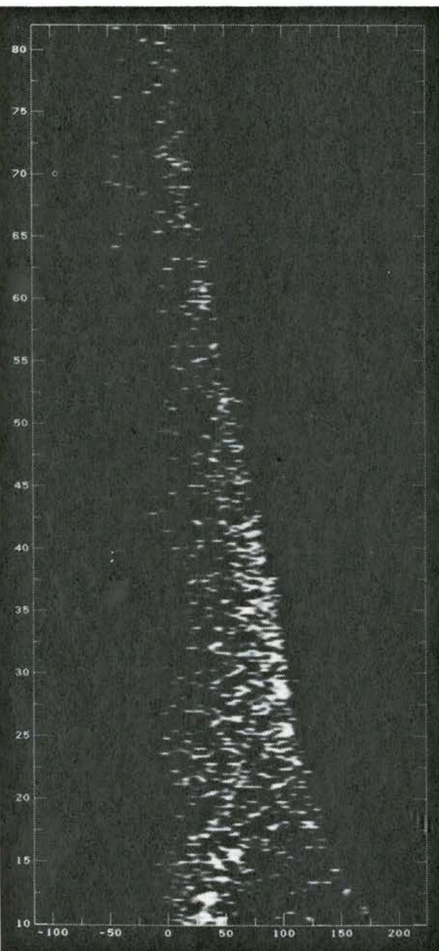
Aunque el modelo reproduce bastante bien muchas de las observaciones, se requirió cierto número de modificaciones para alcanzar su presente éxito. Algunas de las hipótesis iniciales demostraron ser incorrectas, pero se pudieron realizar ajustes razonables para poner al modelo de acuerdo con las observaciones. Por ejemplo, para reproducir la dispersión de las observaciones tuvimos que suponer que la componente aleatoria de la velocidad de una de las nubes respecto a otra tiene una amplitud de unos cuatro kilómetros por segundo. Esta alta velocidad relativa, combinada con una separación media de las nubes aproximadamente sólo unas 10 veces el diámetro medio de las nubes, muestra que el choque y subsiguiente soldadura de las nubes para formar grandes aglomerados no son fenómenos raros. El descubrimiento conduce a una conclusión interesante: las nubes gigantes que probablemente se engarzan de esta manera tienen masas gaseosas totales del orden de 100.000 veces la masa del Sol, que es una estrella bastante típica. En resumen, esos objetos recientemente descubiertos son los objetos más masivos de la galaxia.

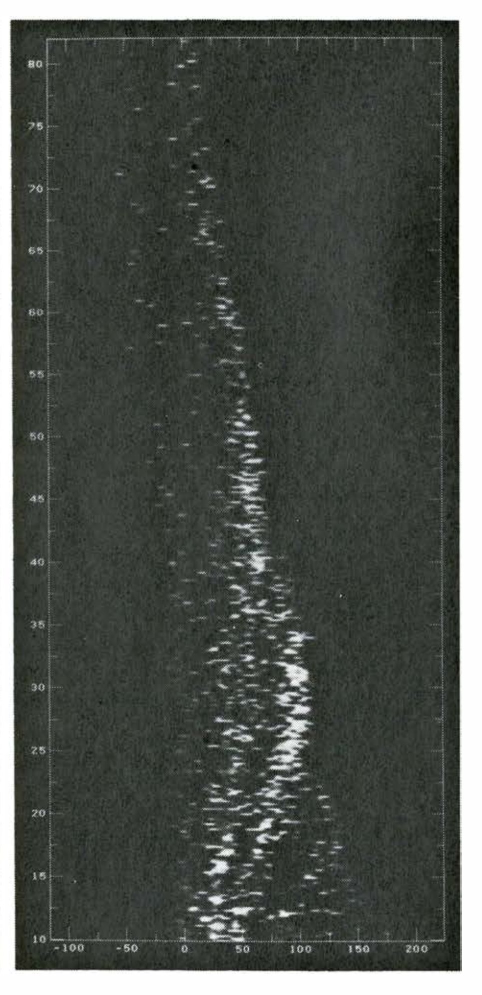
Nuestro modelo final contiene varios cientos de miles de nubes, cada una con un diámetro representativo de 50 años-luz. La separación media de las nubes en la región más densa de la parte media del anillo molecular es de unos 400 años-luz, distancia que crece hacia los bordes del anillo. En realidad, se sabe que la mayor parte de las nubes tienen tamaños que varían desde 15 hasta 100 años-luz, pero la incorporación de un espectro de tamaños de nubes en el modelo probablemente no afectaría los resultados de manera substancial. La masa total de la materia de la nube molecular en la galaxia se estima en varios miles de millones de masas solares.

Queda todavía un serio fallo en el modelo. Hemos supuesto que las nubes están distribuidas por todo el anillo de monóxido de carbono de manera más o menos uniforme en azimut. Cuando se comparan los espectrogramas reales de monóxido de carbono con los producidos por el ordenador a partir del modelo, sin embargo, los datos reales muestran un grado de agrupamiento mayor que las nubes producidas por el ordenador, que están distribuidas más uniformemente a lo largo de la región velocidad-longitud.

La explicación más probable es que las nubes oscuras se encuentran a lo largo de fragmentos de los brazos espirales de la galaxia, y como tales no están distribuidas de manera azimutalmente simétrica en torno al centro de la galaxia.







COMPARACION de tres grupos de espectogramas. El diagrama velocidad-longitud a la izquierda se compone de un gran número de radioespectrogramas de monóxido de carbono galáctico. Los otros dos diagramas se generaron mediante un ordenador y están basados en los modelos de Monte Carlo de los autores referentes a la distribución de las nubes de monóxido de carbono en la galaxia. Las nubes representadas por el diagrama del centro están distribuidas al azar, pero básicamente de manera uniforme. La concordancia entre muchos aspectos de las observaciones y las predicciones del modelo indican que los valores de variables tales como el número de nubes y su separación se escogieron correctamente. El defecto más conspicuo de este modelo es su incapacidad para dar cuenta del agrupamiento de orden superior. En el modelo representado a la derecha, calculado por los autores en colaboración con W. W. Roberts, la distribución y movimientos de las nubes se suponen estar influenciadas por los efectos gravitatorios de los brazos espirales. El hecho de que este modelo satisface las observaciones mejor que el del centro sugiere que, en nuestra galaxia, las regiones generadoras de estrellas describen una configuración espiral, situación observada en las galaxias espirales externas. (Las fotos se tomaron en Kitt Peak.)

W. W. Roberts, de la Universidad de Virginia, ha elaborado descripciones teóricas de los brazos espirales en términos de ondas de choque inducidas gravitatoriamente en el gas interestelar de las galaxias. El concepto central de esta teoría de ondas de densidad es que el gas interestelar y las estrellas se mueven a través de las ondas espirales en su rotación en torno al centro galáctico. Siendo de gran masa, las estrellas pasan a través de las ondas sin ser afectadas, pero el gas interestelar se comprime, formando en el proceso nubes oscuras. Las nubes moleculares serían tan densas en sí que responderían a subsiguientes perturbaciones de la onda de densidad, comportándose más como estrellas que como gas interestelar. Ahora bien, como se cree que la vida de las nubes oscuras es corta comparada con el tiempo que necesitan para emigrar de una onda espiral a otra, un observador vería las nubes oscuras ligeramente corriente abajo de la onda espiral, delineando así la onda. Presumiblemente, las nubes serían visibles por otra razón: se formarían estrellas en ellas, de modo que la onda espiral estaría también marcada por jóvenes estrellas muy luminosas.

La incorporación del concepto de ondas espirales en nuestro modelo de cálculo mejora la concordancia con las observaciones. La ilustración de la página anterior muestra las observaciones, una junto a otra, de un modelo de distribución uniforme y de un modelo de distribución de nubes con ondas de densidad calculado en colaboración con Roberts. La larga configuración a la velocidad de 50 kilómetros por segundo observada entre las longitudes de 50 y 25 grados aparece ahora en el modelo de onda de densidad, pero el mayor agrupamiento de los datos todavía no es reproducido por el modelo. Además, el intervalo de velocidades de las características producidas por las ondas en el modelo es bastante estrecho comparado con el de las observaciones.

Una posible explicación de la incapacidad del modelo para reproducir el alto orden de agrupamiento de los datos es que hay regiones en la galaxia, quizás a lo largo de los brazos espirales, donde las condiciones para la formación de estrellas son desusadamente favorables. En esas regiones hay un exceso de nubes oscuras. Las fotografías de las galaxias espirales apoyan esta interpretación. Típicamente, muestran regiones definidas de alta luminosidad en lugares donde estarían situadas las nubes oscuras. Como éstas tienen larga vida, son grandes y se mueven cada una respecto a las otras, a veces llegan a chocar y de esa manera se reúnen dentro de ciertos límites.

Aunque todavía no hemos percibido los detalles del cuadro general de la galaxia de la Vía Lactea, las observaciones han establecido importantes limitaciones para la formación de estrellas en sus regiones internas y ópticamente inaccesibles. En particular, han aislado las regiones de formación de estrellas.

Comparando nuestras estimas de la relación de la masa en las estrellas a la del gas interestelar, encontramos que entre 12.000 y 24.000 años-luz del centro galáctico, el valor de la relación se mantiene notablemente constante en 25 a uno. ¿Cómo se explica esta constancia? ¿Ha alcanzado la formación de estrellas su punto de saturación en esta parte de la galaxia y está ahora limitada por la pequeña cantidad de gas disponible? De acuerdo con la teoría de la onda de densidad para la estructura galáctica, el valor de la compresión debe aumentar en el sentido hacia el núcleo galáctico. Podría suponerse que la velocidad de formación de estrellas aumentaría de acuerdo con ello. Sin embargo, las únicas indicaciones existentes, aun siendo provisionales, sugieren que la relación de estrellas a gas en esta región es constante.

Otro descubrimiento curioso es que el espesor de la galaxia, al menos el espesor medido por la capa de nubes oscuras, no es muy grande: alrededor de 300 años-luz. Además, el espesor permanece casi constante en toda la extensión del anillo de monóxido de carbono. El espesor viene determinado por el equilibrio entre la compresión gravitatoria y las fuerzas dinámicas y magnéticas de expansión. Sin embargo, en un amplio intervalo de condiciones referentes a nubes y movimiento, parece que el espesor de la capa de monóxido de carbono permanece constante. ¿A qué se debe esta uniformidad?

Como hemos indicado, nuestras medidas sitúan el sistema solar cerca del borde de la región visible de la galaxia, en el borde externo de la región más activa en cuanto a formación de estrellas. Las nubes oscuras que se ven ópticamente desde esta perspectiva representan solamente un pequeño número de las existentes en la parte interna de la galaxia. Si el sistema solar estuviera más próximo al núcleo galáctico, la vida en la Tierra podría estar en peligro porque la probabilidad de que el sistema solar chocase con una nube activa habría aumentado. Si tal choque se produjese, el aumento de la extinción de radiación solar entre el Sol y la Tierra reduciría substancialmente la cantidad de luz solar recibida en la Tierra. De hecho, algunos astrónomos creen que los periodos glaciales en la Tierra se produjeron como consecuencia de encuentros casuales con nubes de polvo interestelar. Si este fuese el caso, sería irónico que las nubes de polvo requeridas para la formación de estrellas y planetas, lo que es presumiblemente una condición previa para la evolución de la vida, fuesen también capaces de producir la extinción de la vida.

¿Existe vida en la Tierra porque el sistema solar está fuera de la principal región polvorienta en la galaxia, o a pesar de ello? La velocidad de formación de estrellas parece ser mucho más alta en la región interna de la galaxia, de modo que uno podría esperar que allí sería mayor la probabilidad de que surgiera la vida. Las probabilidades de choque con una nube de polvo, sin embargo, también son mayores. El problema de cuál de los dos efectos domina y si tal efecto es crucial es objeto de especulación.

La nueva herramienta de la radioespectroscopía de monóxido de carbono, que nosotros y otros estamos empleando para explorar la galaxia, es en algunos aspectos todavía primitiva. La radioastronomía de ondas milimétricas tiene sólo diez años de vida. La mayor parte de los datos referentes al monóxido de carbono han sido recogidos por radiotelescopios en los Estados Unidos. El paso siguiente será hacer observaciones de la otra mitad de la galaxia, estudio que requiere un telescopio de ondas milimétricas situado a una latitud más austral. Se necesitan también receptores más sensibles. Los resultados del trabajo realizado hasta ahora han hecho ya variar radicalmente las ideas sobre la formación de estrellas y la evolución galáctica. Como resultado de los descubrimientos en astronomía de ondas milimétricas ha surgido el nuevo campo de la astroquímica. Se pueden ahora investigar procesos químicos que afectan a una gran variedad de moléculas a las temperaturas y densidades extraordinariamente bajas del gas interestelar, condiciones que no se pueden reproducir en el laboratorio.

Varios países están ahora construyendo grandes radiotelescopios de ondas milimétricas. Los Estados Unidos proyectan levantar un observatorio nacional de esta clase en Hawai para continuar en el camino señalado por los descubrimientos realizados con el telescopio de 11 metros del Observatorio Nacional de Kitt Peak. Las observaciones futuras del monóxido de carbono y de otras moléculas interestelares conducirán de seguro a un más profundo conocimiento de la manera como se forman las estrellas y evolucionan las galaxias.

El spin del protón

Cuando dos protones chocan, los productos que resultan dependen de la manera como las partículas giran. El efecto visto en colisiones violentas sugiere la existencia de objetos en el protón que giran muy deprisa

Alan D. Krisch

Todas las partículas fundamentales de la materia –el protón, el neutrón y el electrón– parecen estar girando perpetuamente. El giro se parece mucho al de una peonza, pero con una diferencia significativa: no hace falta imprimir a las partículas el movimiento giratorio. La rotación es una de sus propiedades intrínsecas. Cada partícula tiene un momento angular de spin fijo, de la misma manera que tiene una masa y una carga eléctrica definidas.

Cuando dos partículas con spin chocan, los productos del choque a menudo dependen de cómo están girando aquéllas. Por ejemplo, la trayectoria que sigue un protón después de una colisión puede estar afectada por el spin del protón, de la misma manera que un efecto aplicado a una bola de billar puede alterar la trayectoria de la bola. Durante mucho tiempo se ha pensado, sin embargo, que la influencia del spin debe decrecer a medida que aumenta la energía de la colisión. La razón de esta hipótesis es simple: la energía asociada con el spin de un protón es constante, y es así una fracción cada vez más pequeña de la energía total cuando la colisión se hace más violenta. Para una energía de colisión suficientemente alta no debe importar el que los dos protones que chocan estén girando de la misma manera o en direcciones opuestas.

Sólo en años recentísimos se han desarrollado técnicas experimentales para probar esta hipótesis. Ha resultado ser bastante errónea. La influencia delspin no disminuye a medida que la energía de la colisión aumenta: por el contrario, el spin parece hacerse más importante cuando la colisión es más violenta. Un conjunto reciente de experimentos ha mostrado que la probabilidad de que dos protones reboten violentamente es mucho mayor para protones que giren en la misma dirección, que para protones que giren en direcciones opuestas. Los protones con spines opuestos a menudo parecen pasar de largo, sin interaccionar en absoluto.

La interpretación de estos experimentos no está clara todavía. Parecen indicar que dentro del protón hay algunos objetos más pequeños que llevan la mayor parte del momento angular de spin de la partícula. Estos objetos deben estar girando muy rápidamente. Algunos años antes de que se conocieran estos resultados se había propuesto ya que el protón tenía una estructura interna. Un modelo sugería que cada protón poseía un corazón ("core") denso y pequeño. Una teoría popular hoy día supone que cada protón está formado por tres de los pequeños entes llamados quarks. Se han acumulado razones independientes en favor de la hipótesis de los quarks. Por otro lado, experimentos recientes sugieren que los constituyentes internos del protón tienen propiedades asociadas con su spin que son algo distintas de las propiedades predichas por la hipótesis de los quarks. No es fácil reconciliar estas diferencias.

El momento angular de un objeto es proporcional a su velocidad de rotación y está también condicionado por la distribución de su masa. Una partícula elemental puede tener dos clases de momento angular: el momento angular orbital y el de spin. El momento angular orbital se visualiza mejor en el modelo planetario del átomo propuesto por Niels Bohr en 1913. El momento angular orbital de un electrón en un tal átomo es proporcional a la velocidad con que gira alrededor del núcleo, al radio de su órbita y a su masa. El momento angular orbital aparece en cualquier ligera colisión entre partículas; aun cuando las partículas no describan un círculo completo, realizan un breve giro alrededor de su centro de masas común.

El momento angular de spin mide la rotación de cada partícula alrededor de su propio eje interno. En mecánica cuántica el spin se diferencia de una manera fundamental del momento angular orbital. Una partícula puede ganar o perder momento angular orbital en función de sus circunstancias, como cuando un electrón en un átomo salta desde un nivel energético a otro y, por tanto, desde una órbita a otra. En tanto que el momento angular de spin es una propiedad fija de cada partícula. La magnitud del spin puede cambiarse solamente alterando la identidad de la propia partícula.

El momento angular de spin se describe corrientemente como un vector, una cantidad que tiene a la vez módulo y dirección. El vector de spin puede representarse como una flecha paralela al eje de rotación y de longitud proporcional a la magnitud del spin. La dirección de la flecha se define por una convención arbitraria llamada regla de la mano derecha. Si los dedos de la mano derecha se cierran alrededor de la partícula en la dirección en que ésta gira, entonces el pulgar indica la dirección del vector de spin. Según esta convención, el momento angular de spin de la tierra podría representarse por un vector en el Polo Norte dirigido hacia arriba.

El concepto de spin se introdujo en la física de las partículas elementales en 1925 por Samuel A. Goudsmit y George E. Uhlenbeck. A fin de explicar el desdoblamiento de ciertas líneas en los espectros de los átomos, propusieron que cada electrón tenía un spin intrínseco. Su propuesta fue hecha en el contexto de la teoría mecánico-cuántica del átomo, que entonces estaba empezando a formularse. Supusieron que tanto el módulo como la dirección del momento angular de spin podía tomar sólo valores discretos.

El módulo del vector de spin se mide en unidades de $\hbar$, un símbolo que se lee "h barra" y que es igual a la constante de Planck h dividida por 2π; $\hbar$ es la unidad más fundamental de momento angular. En el modelo atómico de Bohr de 1913, el momento angular orbital de un electrón sólo podía tomar ciertos valores, tales como $0, 1\hbar, 2\hbar, 3\hbar$, etc. Goudsmit y Uhlenbeck sugirieron que el momento

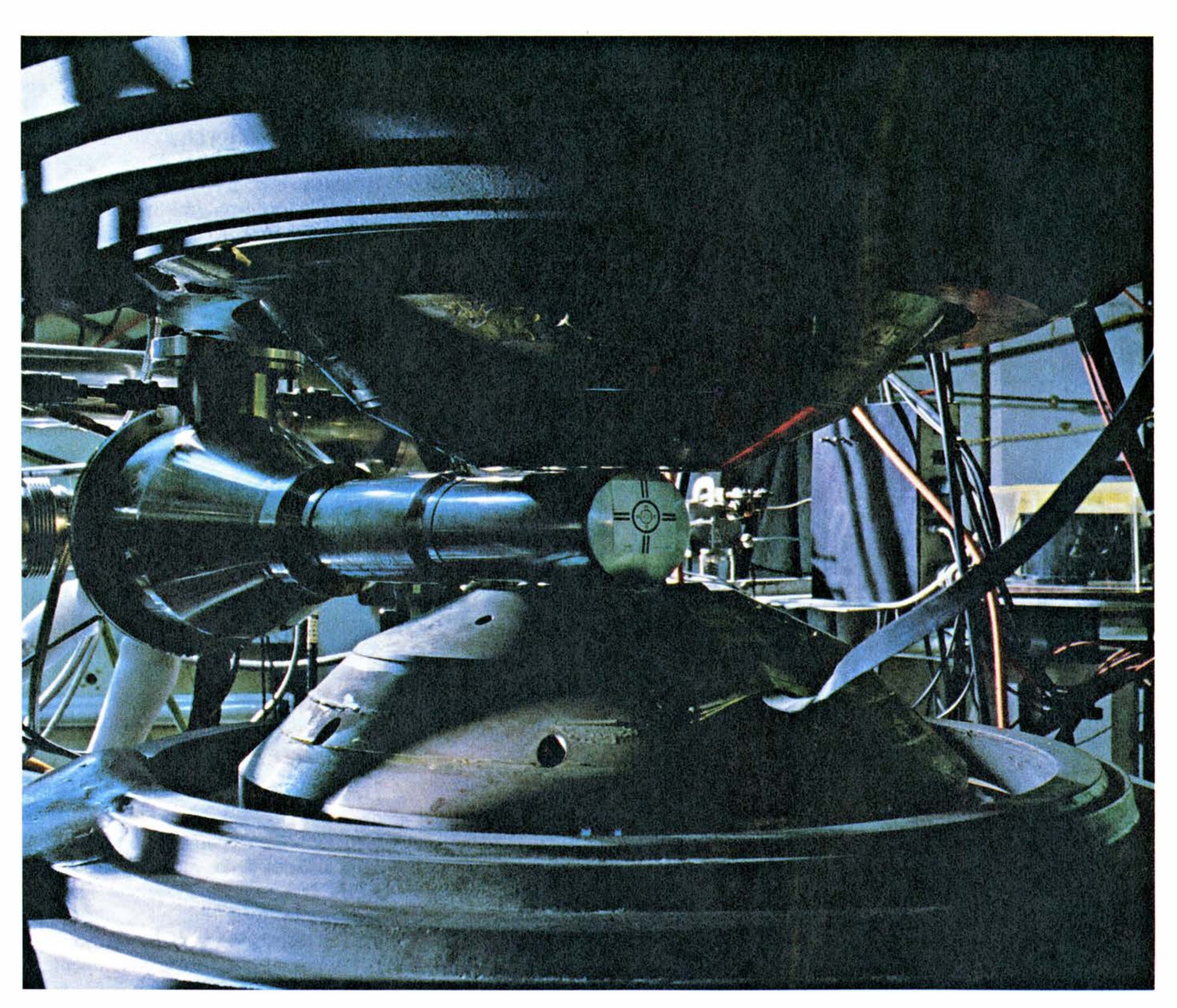
angular de spin de un electrón es exactamente $1/2\ \hbar$, un resultado que desde entonces ha sido confirmado por numerosos experimentos. Además, se ha encontrado que el protón y el neutrón tienen el mismo spin que el electrón.

De la misma manera que el módulo del vector de spin puede tomar únicamente algunos pocos valores, también sólo puede tomar unas pocas direcciones. El número de direcciones permitidas se relaciona con el módulo: para el protón, el neutrón y el electrón (y para todas las otras partículas de spin $1/2\ \hbar$) el vector sólo puede tener dos direcciones, que son siempre diametralmente opuestas. Las dos direcciones pueden definirse respecto de cualquier sistema de referencia, pero es costumbre referirse a ellas simplemente como "arriba" ("up") y "abajo" ("down").

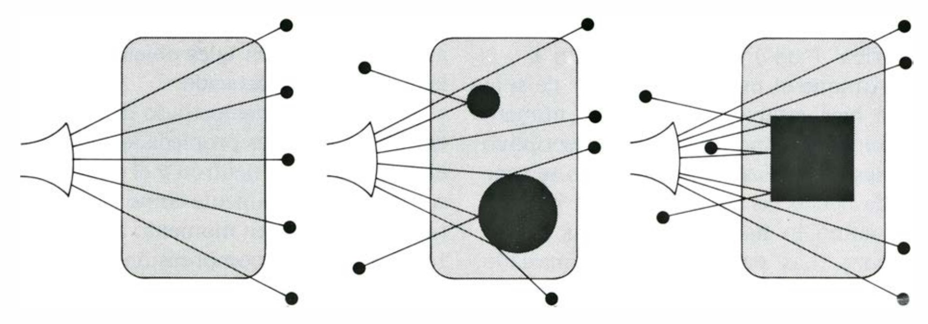
Aunque el momento angular de spin está bien establecido como un número cuántico y como un atributo descriptivo de las partículas elementales, no se sabe si de hecho las partículas están girando tal como lo hacen los planetas o una peonza. Las partículas son demasiado pequeñas para que se pueda observar directamente su rotación. El protón y el neutrón tienen un radio aproximado de un fermi, una unidad de longitud igual a 10^{-13} centímetros. Es posible imaginar, en principio, que una tal partícula pueda estar girando como una peonza ordinaria. El electrón y el neutrino, sin embargo, tienen el mismo spin de $1/2\hbar$ y son aún más pequeños; realmente pueden incluso ser partículas puntuales, sin tamaño. El sentido de un momento angular de spin en tales objetos se halla lejos de estar esclarecido.

Lo que se ha demostrado desde 1925 es que, en todas las propiedades observables, el protón, el neutrón y el electrón se comportan exactamente como si estuvieran girando con un momento angular de $1/2\hbar$. Cualquier comprensión más profunda debe aguardar a experimentos en los que los efectos del momento angular de spin se analicen directamente. Los experimentos que describiré aquí pueden representar el primer paso de un tal programa.

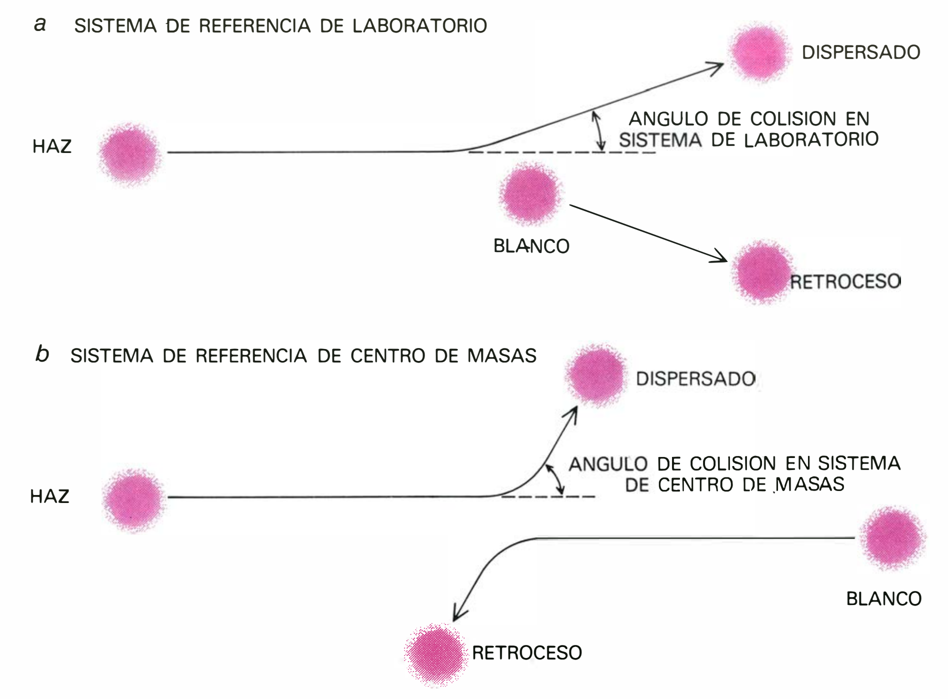
El experimento prototipo de la física de las partículas elementales se hizo en 1911 en el laboratorio de Ernest Rutherford. Se dirigió un haz de partículas alfa, o núcleos de helio, contra una delgada



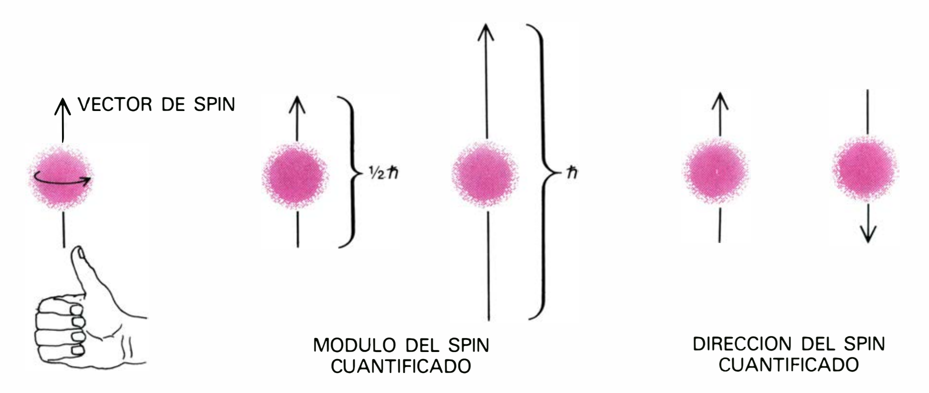
INFLUENCIA DEL SPIN en las colisiones entre protones; se estudia en el Argonne National Laboratory con este aparato. El cilindro metálico del centro encierra un blanco en el que los protones están orientados de manera que la mayoría de ellos giran en la misma dirección; se dice entonces que los protones están polarizados. El blanco es golpeado por un haz de protones que también están polarizados. El haz se acelera mediante el Sincrotrón de Gradiente Cero (Zero Gradient Synchrotron, ZGS), de Argonne, y pasa a través del blanco de izquierda a derecha. El blanco está colocado entre los polos de un imán de 25 toneladas, que forma parte del sistema que alinea los spines de los protones del blanco. El alargamiento cónico del final del cilindro contiene el helio líquido refrigerante. La diana cerca del extremo del cilindro es una marca indicadora que ayuda a dirigir el haz de protones.



LOS EXPERIMENTOS DE COLISION analizan la estructura del protón por un método que se ilustra aquí mediante una analogía. El protón puede imaginarse como algo parecido a una bala de algodón cuyo interior es inaccesible a una observación directa. Un experimentador puede estudiar el interior disparando muchos proyectiles a través de la bala. Si todos los proyectiles la atraviesan sin desviarse, probablemente no habrá nada en la bala. Si observamos que algunos proyectiles se desvían, ello indicará que hay un objeto duro dentro. El número de desviaciones señalará el tamaño del objeto y los ángulos con que los proyectiles rebotan proporcionará información acerca de su forma. En los experimentos de Argonne los "proyectiles" empleados como sondas del protón son, a su vez, protones.



UNA COLISION ELASTICA ocurre cuando dos protones chocan y rebotan, llevándose toda la energía de la colisión, sin crear nuevas partículas. En el sistema de referencia del laboratorio (*a*) un suceso de colisión consiste en un protón acelerado que golpea a otro en reposo. Las trayectorias son más simétricas en otro sistema de coordenadas (*b*), que se mueve con el centro de masas del sistema de dos partículas. En este sistema de referencia, los protones chocan frontalmente con igual velocidad. El ángulo de colisión en el sistema de coordenadas del centro de masas mide la violencia de la colisión.



EL MOMENTO ANGULAR DE SPIN se representa como un vector, es decir, una cantidad con módulo y dirección. El vector de spin es paralelo al eje de rotación y su dirección se da por la regla de la mano derecha: si los dedos de la mano derecha se curvan en la dirección de la rotación, el pulgar señala la dirección del vector. Para el protón el módulo del vector de spin es $1/2\hbar$, donde $\hbar$ es una unidad de momento angular igual a la constante de Planck dividida por 2π (el electrón y el neutrón tienen este mismo spin); para otras partículas el módulo es distinto, pero es siempre un múltiplo de $1/2\ \hbar$. La dirección del vector también está cuantificada. El spin del protón tiene dos direcciones posibles.

lámina de oro. Las partículas dispersadas en las colisiones con los átomos de oro fueron detectadas observando el centelleo que causaban en una pantalla fluorescente. A partir del número inesperadamente elevado de partículas que fueron desviadas en ángulos grandes, Rutherford dedujo que el átomo de oro tenía un corazón pequeño y denso: el núcleo atómico.

La mayoría de los modernos experimentos de dispersión ("scattering") son de concepción idéntica al de Rutherford: difieren sólo en la escala. El aparato se ha hecho más grande y las partículas estudiadas se han hecho más pequeñas. Los proyectiles empleados hoy día suelen ser protones o electrones acelerados a energías que se miden en miles de millones de electronvolt o GeV; algunas de las máquinas necesarias para alcanzar estas energías tienen una circunferencia de varios kilómetros. Estos aceleradores permiten estudiar, además del núcleo como un todo y los protones y neutrones constituyentes del mismo, la estructura interna del protón y el neutrón.

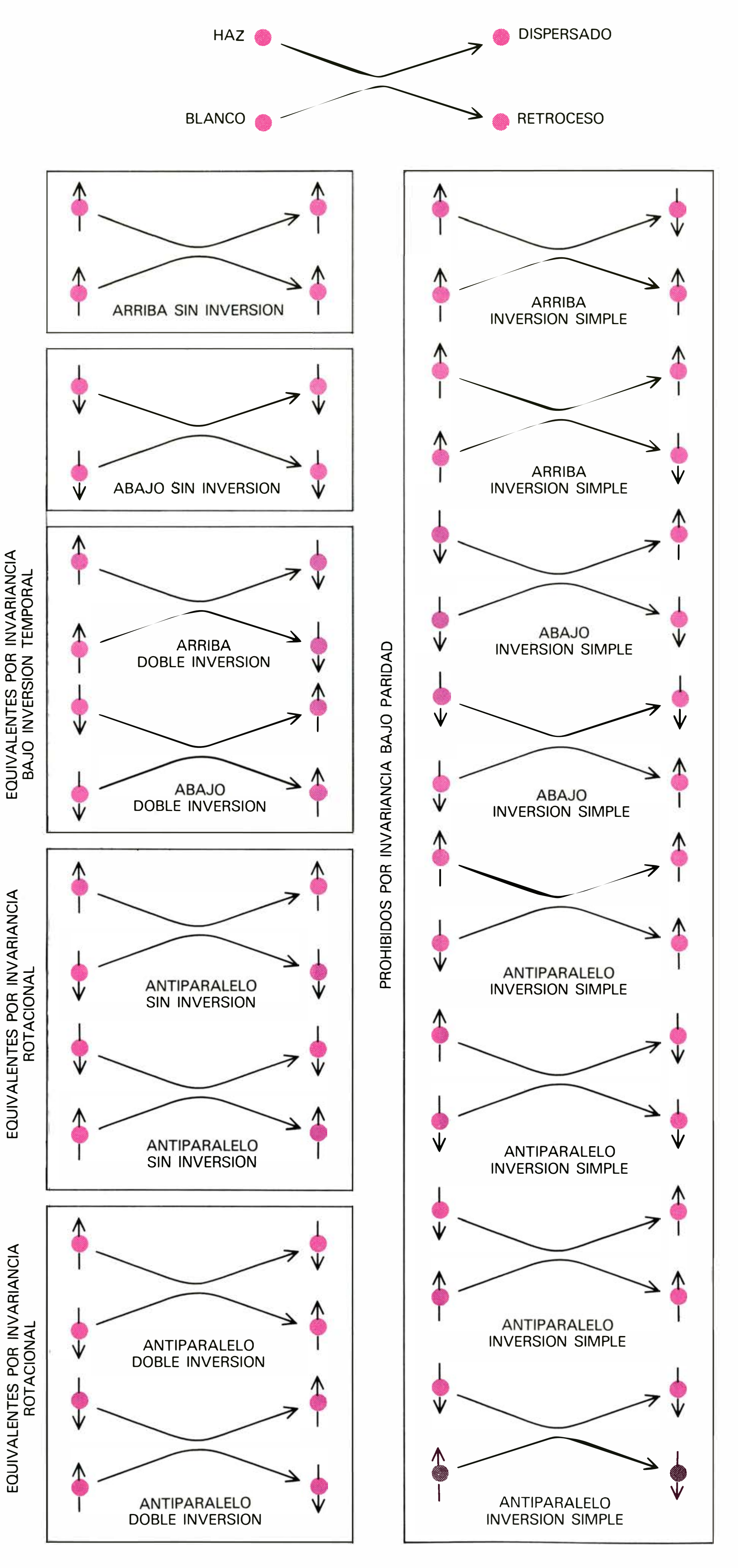
Parece irónico que se dirijan los mayores instrumentos científicos hacia los objetos más pequeños, pero hay una explicación clara. El principio de incertidumbre de la mecánica cuántica, formulado por Werner Heisenberg, establece una relación entre la precisión con que un objeto puede medirse y la energía disponible para la medida. El principio establece que el producto de estas dos cantidades es constante; si la precisión se da en fermi y la energía en GeV, entonces el producto es igual a 0,197. Supongamos que se quiera hacer un estudio del protón con una precisión igual al uno por ciento del radio del protón. Esta precisión es de unos 0,01 fermi, o 10^{-15} centímetros, de manera que la energía necesaria será de 19,7 gigaelectronvolt.

De hecho, la energía exigida a un acelerador en un tal experimento es aún mayor. En mecánica newtoniana la energía cinética de una partícula en movimiento es proporcional al cuadrado de su velocidad. Si la partícula se detuviera completamente en una colisión, cedería toda su energía cinética, pero un protón que choca con otro protón no se para. Por el contrario, el protón dispersado y el de retroceso se llevan entre los dos en su movimiento la mitad de la energía disponible. El mismo efecto puede observarse cuando un automóvil en movimiento choca contra otro aparcado: la colisión es mucho menos violenta que si hubiese sido frontal.

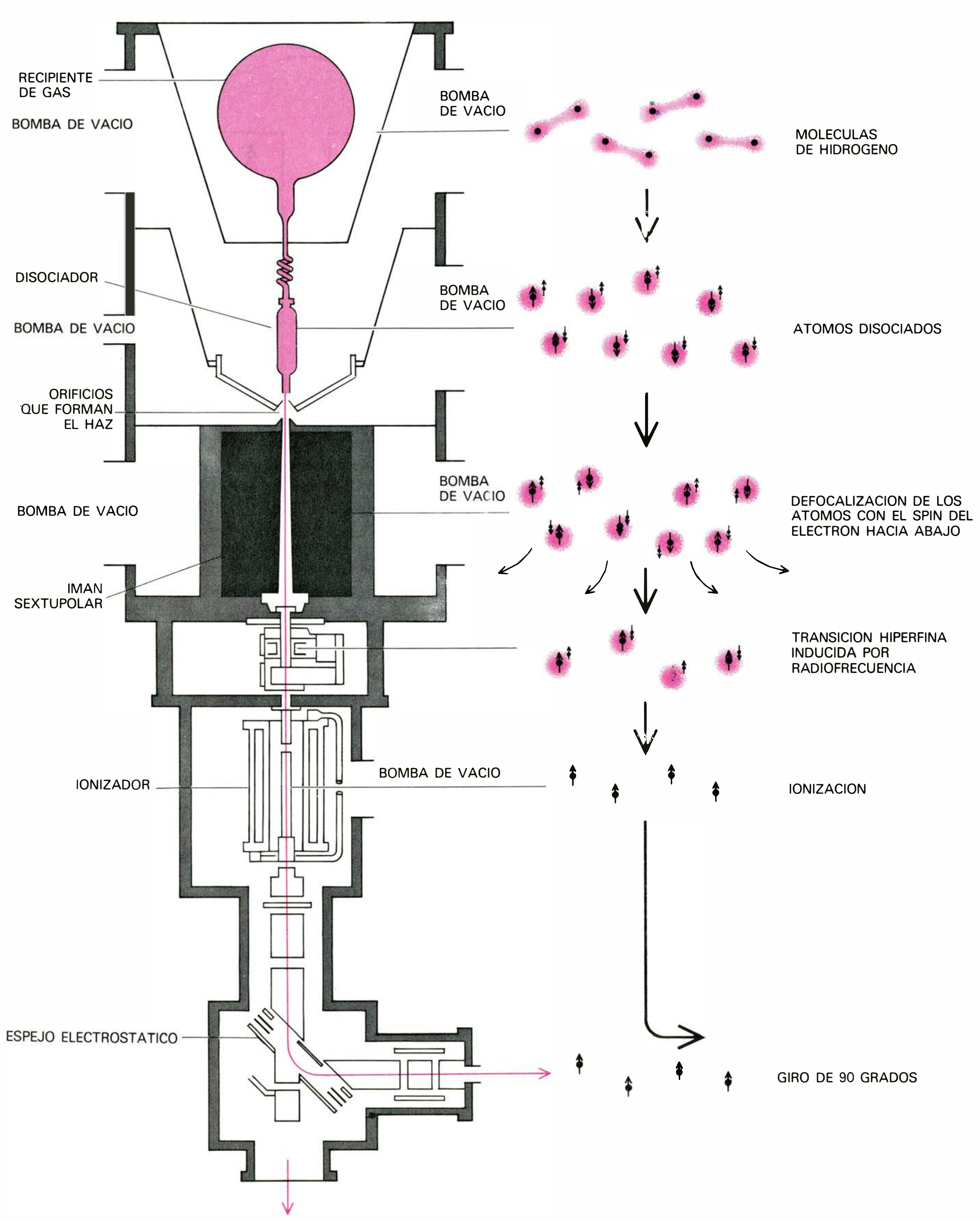
El análisis de estas colisiones puede simplificarse cambiando a otro sistema de referencia. El sistema de coordenadas en que se observa la colisión es el sistema de referencia de laboratorio, en el que el observador y el acelerador están en reposo. Un sistema de coordenadas más natural es el que se mueve con el centro de masas de las dos partículas. En este sistema de referencia cada protón tiene la mitad de la velocidad del haz; como la energía de la partícula es proporcional al cuadrado de su velocidad, cada protón tiene un cuarto de la energía del haz. Es claro así que sólo la mitad de la energía del haz (tal como se mide en el sistema de referencia de laboratorio) está disponible para la colisión.

La teoría de la relatividad especial añade una complicación más. A energías del haz superiores a los 6 GeV, el protón acelerado se mueve ya a una velocidad mayor que el 99 por ciento de la velocidad de la luz; posteriores aumentos en la energía repercuten poco en su velocidad, aumentando en cambio la masa efectiva de la partícula. En consecuencia, la proporción de la energía del haz disponible para la colisión se reduce drásticamente. Una analogía apropiada la constituye el choque de un tren de carga en movimiento con un automóvil en reposo: sólo una pequeña parte de la energía empleada en acelerar el tren queda disponible para la colisión. De hecho, para alcanzar una energía de colisión en el centro de masas de 19,7 GeV, hace falta una energía de haz de unos 800 GeV, aproximadamente el doble de la energía suministrada por el mayor de los aceleradores de partículas que funcionan actualmente.

Otro factor, aparte de la energía del centro de masas, afecta al poder de resolución de un experimento de acelerador: se trata del ángulo que se desvían las partículas. Las colisiones suaves, en las que hay poco cambio en la energía o dirección de las partículas, muestran sólo la interacción de largo alcance entre los protones. Las características más finas de la estructura del protón se observan sólo



LAS SECCIONES EFICACES de spin puro miden la probabilidad de que dos protones colisionen cuando sus vectores de spin apunten en direcciones determinadas antes y después del suceso. Hay 16 posibles configuraciones de spin, pero ocho corresponden a sucesos prohibidos por la exigencia de que la paridad se conserve. Entre las ocho configuraciones restantes hay tres pares formados por secciones eficaces que no pueden distinguirse una de otra a causa de simetrías o principios de invariancia. Cada par puede considerarse como una única sección eficaz, de manera que hay cinco secciones eficaces de spin puro distintas. Los experimentos del laboratorio de Argonne miden diferencias entre estas cinco secciones eficaces de spin.



FUENTE DE PROTONES POLARIZADOS para el Sincrotrón de Gradiente Cero (Zero Gradient Synchrotron). La fuente alinea los spines de los protones polarizando primero los electrones. Las moléculas de hidrógeno se disocian en átomos, que surgen en forma de haz y entran en el campo no uniforme de un imán sextupolar. Todos los átomos que tienen los electrones con spin hacia abajo son dirigidos hacia la periferia del campo y por tanto se eliminan del haz. En los átomos restantes todos los spines de los electrones están hacia arriba y los spines de los protones están orientados al azar, mitad hacia arriba y mitad hacia abajo. Después los átomos se someten a radiación de radiofrecuencia sintonizada de manera que sólo pueda ser absorbida por átomos en que el electrón y el protón tengan spines opuestos. La absorción de la radiación invierte ambos spines, un suceso llamado transición hiperfina. Después de la doble inversión del spin todos los spines de los protones están orientados hacia arriba y los spines de los electrones están orientados al azar (mitad arriba y mitad abajo). Finalmente se arrancan los electrones y el haz de protones polarizados resultante se refleja en un ángulo de 90 grados, de manera que los vectores de spin son perpendiculares a la dirección del movimiento del haz. La conaminación por átomos errantes estropearía la polarización; por consiguiente, vemos que la fuente del haz de protones exige un elaborado sistema físico de bombas de vacío.

en las colisiones más violentas en que los dos protones se separan con ángulos de desviación elevados. Contando el número de protones desviados en cada dirección, puede estimarse el tamaño y la forma de los protones que chocan. De hecho fue a partir de un tal experimento como Rutherford dedujo el tamaño del núcleo.

Lo que se mide de hecho en un experimento de colisión es la cantidad llamada sección eficaz [véase "Colisiones de hadrones a muy alta energía", por Carlos Pajares y Ramón Pascual; INVESTIGACION Y CIENCIA, octubre, 1978]. Como sugiere el nombre, una sección eficaz puede expresarse en unidades de área, pero de hecho es una medida de la probabilidad de que una colisión tenga lugar. Por ejemplo, los experimentos hechos recientemente por mis colegas y yo medían la sección eficaz de colisión elástica cuando dos protones chocan. Se dice que una colisión es elástica si toda la energía del protón incidente se la llevan los dos protones salientes; en los sucesos inelásticos una parte de la energía se emplea en crear nuevas partículas. La sección eficaz se midió para varios ángulos de colisión a una energía de 12 GeV. Los experimentos fueron realizados por un grupo de cerca de una docena de físicos, principalmente de la Universidad de Michigan y del Argonne National Laboratory cerca de Chicago. El acelerador que empleamos fue el Sincrotrón de Gradiente Cero (Zero Gradient Synchrotron, ZGS) de Argonne.

La sección eficaz elástica de colisión protón-protón había sido medida antes muchas veces y a energías mayores que las proporcionadas por el ZGS. Lo que distinguió nuestros experimentos fue el que nosotros conocíamos las direcciones de los vectores de spin del protón incidente y del protón blanco antes de que chocaran. Podíamos medir por tanto una sección eficaz separada para cada combinación de spines. En experimentos anteriores los spines estaban orientados al azar y de los datos registrados no se podía obtener información sobre su influencia.

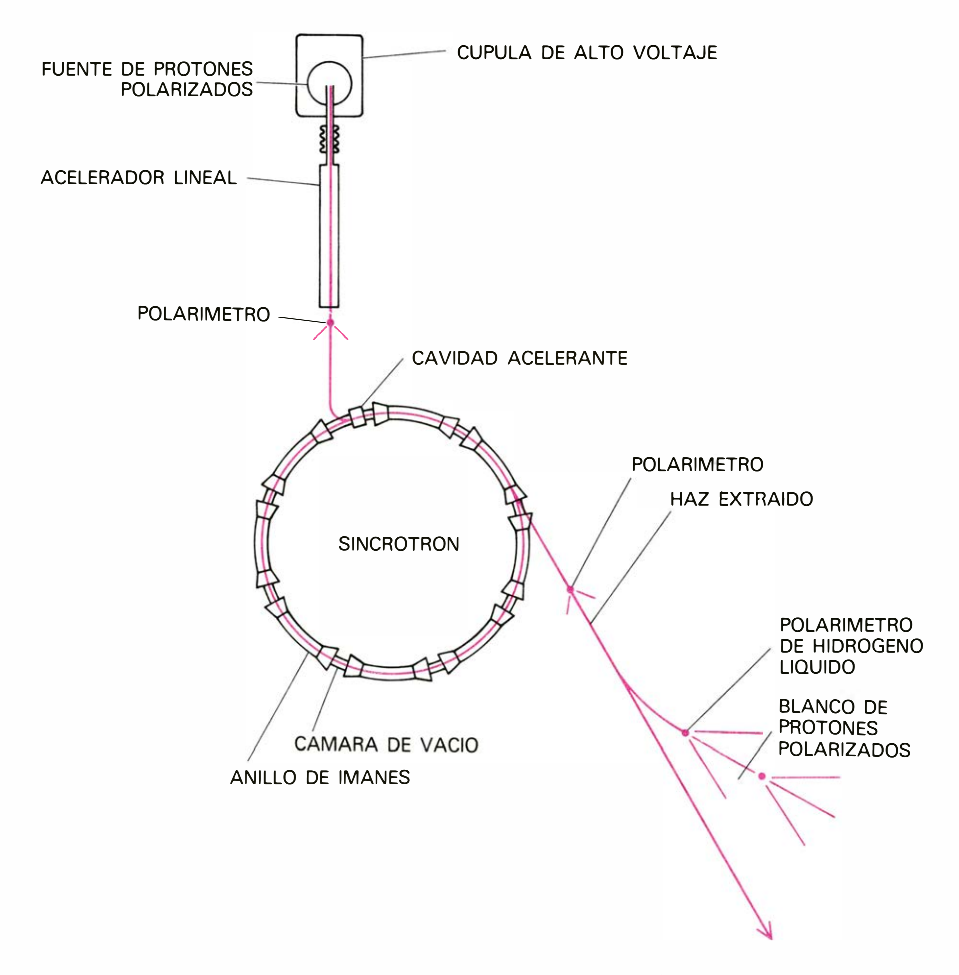
Sólo dos protones participan en un suceso de colisión elástica, pero deben considerarse cuatro spines. Los spines del protón del haz, del blanco, del dispersado y del de retroceso. El protón del haz y el dispersado son de hecho la misma partícula, de la misma manera que lo son el protón del blanco y el de retroceso; sin embargo, sus spines deben determinarse separadamente porque durante la colisión pueden cambiar de dirección, o en otras palabras, hacer una inversión de 180 grados.

Como hay cuatro spines y cada spin puede tener una de dos direcciones posibles, hay 2^4 o 16 combinaciones de spin posibles. Las 16 secciones eficaces correspondientes sin mezcla de spines son las secciones eficaces fundamentales para la colisión elástica protón-protón. Si alguna vez se desarrolla una teoría que describa con precisión la interacción entre protones, deberá predecir estas secciones eficaces de spin puro. Los experimentos que no pueden distinguir los estados de spin del protón miden sólo un complejo promedio de todas las secciones eficaces.

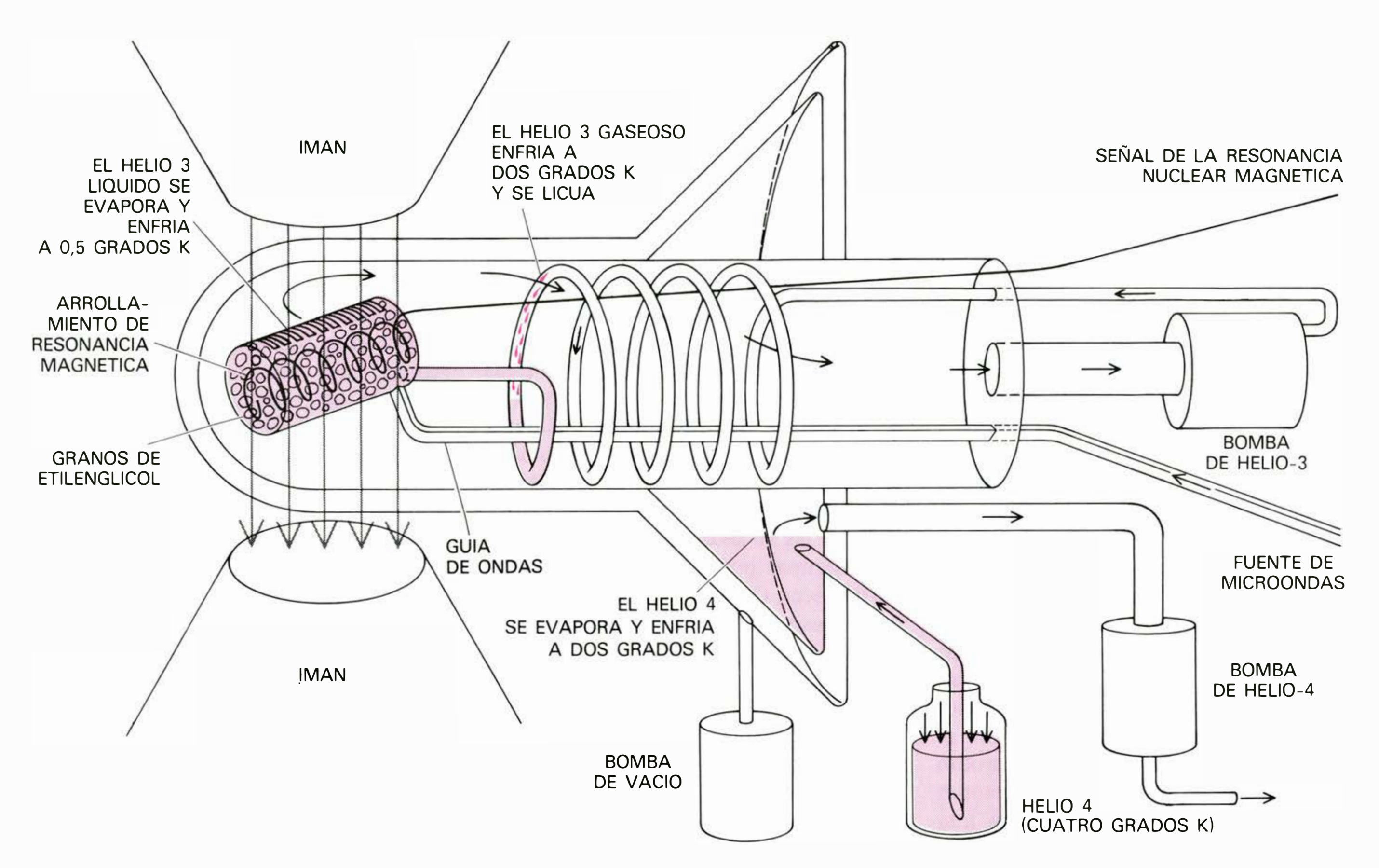
Entre las 16 secciones eficaces de spin puro algunas pueden eliminarse y otras permanecer. Ocho de las secciones eficaces corresponden a sucesos en los que se invierte sólo un spin. Si los dos protones que chocan forman un sistema aislado (como debe ser en un experimento de colisión elástica), todos los sucesos en que hay una sola inversión están prohibidos por una ley de conservación. Las secciones eficaces para estos sucesos deben ser por tanto exactamente cero; tales sucesos no se observan.

No todas las ocho secciones eficaces que subsisten se pueden distinguir experimentalmente entre sí. Dos de ellas son equivalentes por el postulado de invariancia bajo inversión temporal, que establece que cualquier proceso físico microscópico no debe cambiar cuando se invierte la dirección del tiempo. Dos pares más de secciones eficaces son equivalentes por invariancia rotacional: un miembro de cada par se identifica con el otro si giramos todo el laboratorio cabeza abajo.

Al agrupar los sucesos indistinguibles quedan cinco secciones eficaces distintas. En una notación breve conveniente se



SINCROTRON DE GRADIENTE CERO (Zero Gradient Synchrotron). Acelera protones polarizados a una energía de 12.000 millones de electronvolt (GeV). La fuente de protones polarizados está encerrada en una cúpula de alto voltaje; los protones que emergen de la cúpula adquieren una energía de 750.000 electronvolt. Entonces, un acelerador lineal aumenta su energía hasta 50 millones de electronvolt antes de que sean inyectados en el propio sincrotón. En cada paso por la cavidad aceleradora del sincrotón las partículas ganan 12.000 electronvolt; para alcanzar los 12 GeV deben dar la vuelta al anillo alrededor de un millón de veces, lo que requiere algo menos de un segundo. El acelerador se llama sincrotrón de gradiente cero porque los campos de los imanes del anillo son uniformes, lo que ayuda a mantener la polarización del haz. Cuando las partículas alcanzan los 12 GeV, se extraen del anillo y se dirigen por imanes de control hasta el blanco. La polarización se controla antes de la inyección y tras la extracción mediante polarímetros que miden la asimetría izquierda-derecha en sucesos de colisión. La asimetría está relacionada con la orientación del spin y con la polarización del haz.



EL BLANCO DE PROTONES POLARIZADOS se enfría a una temperatura de 0,5 grados Kelvin, se coloca en un campo magnético de 25.000 gauss y se expone a radiación de microondas de una frecuencia de 70 gigahertz. La refrigeración está proporcionada por dos isótopos del helio. El isótopo más común helio 4 enfría el conjunto del aparato a 2 grados Kelvin, y el mayor enfriamiento del propio blanco lo da el isótopo ligero helio 3. Ambos sistemas de refrigeración emplean bombas de vacío para reducir la temperatura de los líquidos por aumento de su evaporación. La energía de microondas es llevada hasta el blanco por una guía de ondas. Un arrollamiento interior al recipiente del blanco mide la polarización del spin de los protones por el método llamado resonancia magnética nuclear. El material del blanco es etilenglicol dopado con cromo, pero sólo se polarizan los protones de ciertos átomos de hidrógeno. A la temperatura del blanco, el etilenglicol es un sólido y se encuentra formando pequeños granos.

las conoce como arriba sin inversión, abajo sin inversión, paralela con doble inversión, antiparalela sin inversión y antiparalela con doble inversión [*véase la ilustración de la página 25*]. Puede parecer que las dos primeras secciones eficaces, que representan respectivamente sucesos con los cuatro spines hacia arriba y los cuatro spines hacia abajo, podrían también relacionarse por invariancia rotacional. De hecho estas secciones eficaces, aunque muy parecidas, son distintas. Debido a que el momento angular de spin produce consecuencias más o menos similares a las observadas en una bola de billar con efecto, los sucesos de colisión con spines arriba sin inversión y abajo sin inversión son ambos asimétricos. Al girar el laboratorio cabeza abajo no se transforman uno en otro.

En la práctica no es siempre necesario distinguir siquiera estas cinco secciones eficaces. La distinción más importante aparentemente está entre los sucesos en que el protón del haz y el protón del blanco están girando de la misma manera y aquellos en que sus giros son opuestos. Estas dos clases de sucesos se denominan convenientemente mediante las palabras paralelo y antiparalelo. Muchos de los resultados que presentaré aquí sólo observan esta distinción.

Una investigación experimental acerca de la influencia del spin en las colisiones protón-protón exige a la vez un haz y un blanco cuyos protones tengan todos, o casi todos, sus spines alineados de la misma manera. Se dice entonces que el haz y el blanco están polarizados. (El uso de este término de origen óptico está totalmente legitimado, ya que la polarización de la luz consiste en alinear los spines de los fotones.) Por ahora el Zero Gradient Synchrotron de Argonne es el único acelerador del mundo que dispone a la vez de un haz de protones polarizados de alta energía y de un blanco de protones polarizados.

La única manera práctica de dominar una partícula y controlar la dirección de su spin es mediante el momento magnético de la partícula, que está estrechamente relacionado con su spin y su carga eléctrica. En un campo magnético externo, el momento magnético del protón da lugar a una pequeña fuerza que tiende a alinear el vector de spin a lo largo de la dirección del campo. Así, un método directo de polarizar protones consiste en la aplicación bruta de un campo magnético fuerte, pero éste no es el mejor medio. El momento magnético del protón es excesivamente pequeño y el campo magnético necesario para polarizar una fracción elevada de los protones de una muestra de materia sería impracticablemente grande.

En el ZGS se emplea un procedimiento más complicado, pero también más eficiente, en una ingeniosa fuente de protones polarizados construida por la ANAC Corporation de New Zealand. El principio básico de la operación consiste en actuar sobre el momento magnético del electrón, que es unas 660 veces mayor que el del protón. Se polarizan los electrones y entonces la alineación de sus spines se transfiere a los protones.

Al principio, los protones están formando los núcleos de moléculas de gas hidrógeno (H_2), que se disocian en átomos. En cada átomo tanto el protón como el electrón tienen un spin con dos orientaciones posibles, y así hay cuatro

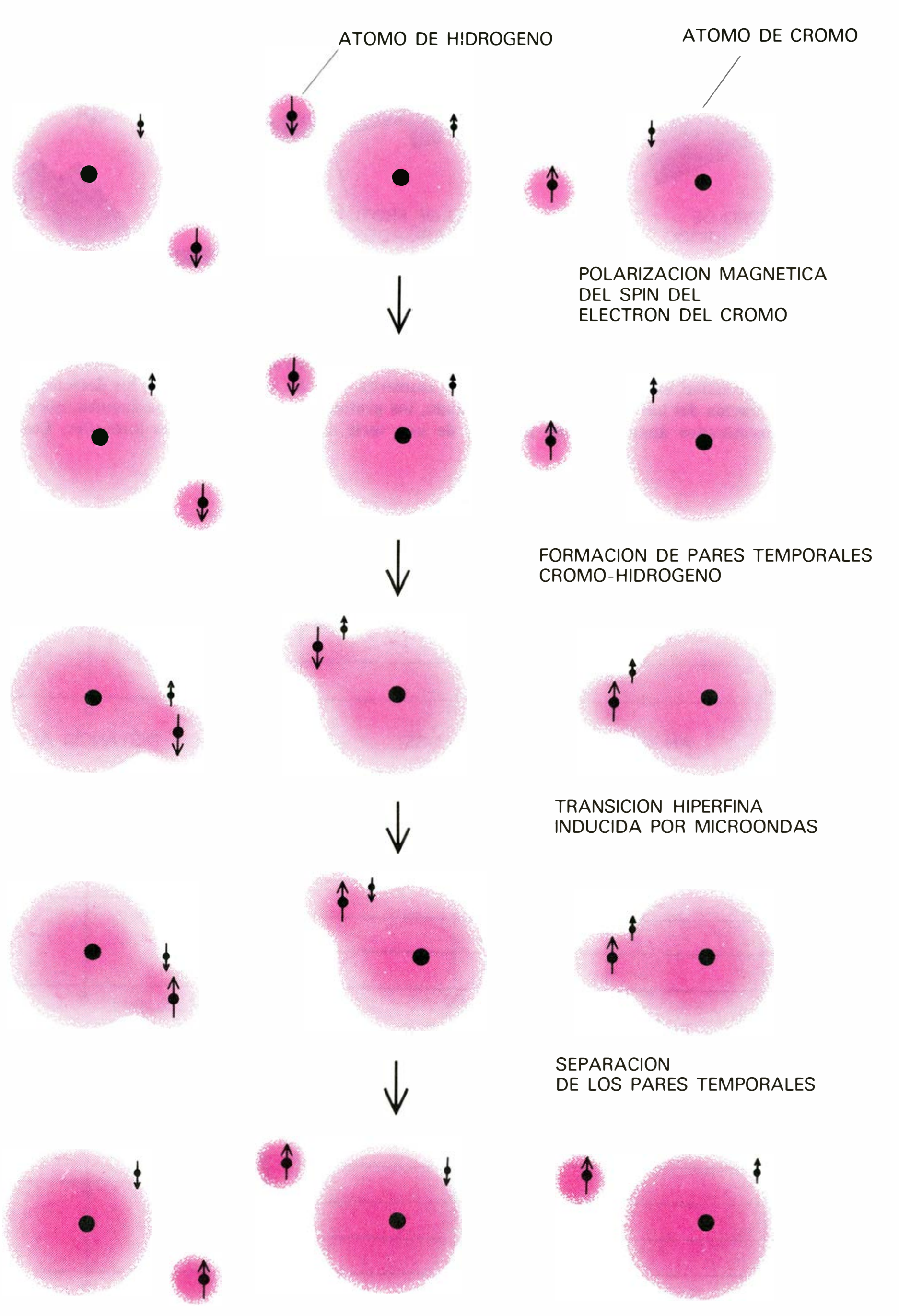
posibles combinaciones de spin. Los átomos emergen por una rendija formando un haz de baja energía, pasando entonces por un campo intenso y no uniforme creado por un imán sextupolar, o sextupolo. El campo tiene poco efecto directo sobre los spines de los protones, pero discrimina muy fuertemente entre los electrones de spin arriba y de spin abajo. Los átomos con electrones de spin arriba son repelidos por el campo y se focalizan por tanto a lo largo del eje central del imán, donde el campo es mínimo. Los átomos con electrones de spin abajo son arrastrados a la región de campo más fuerte de la periferia y quedan por tanto excluidos del haz.

Saliendo del imán sextupolar queda un haz de átomos con los spines de los electrones completamente polarizados y los spines de los protones completamente no polarizados, es decir, la mitad arriba y la otra mitad abajo. El paso siguiente polariza los protones y despolariza los electrones. Ello se logra invirtiendo a la vez los spines de los electrones y los protones en una población de átomos, aquella en que los spines son antiparalelos. La inversión selectiva de los spines se logra aplicando un campo magnético uniforme e irradiando los átomos a una frecuencia de 1,5 gigahertz (mil millones de ciclos por segundo). Esta frecuencia es absorbida sólo por aquellos átomos en que el electrón tiene spin arriba y el protón tiene spin abajo, e induce una doble inversión de spin llamada transición hiperfina. Después de la transición, el electrón tiene spin abajo y el protón spin arriba. Los únicos otros átomos presentes tienen el electrón y el protón con spin arriba, y no quedan afectados por la radiación de esta frecuencia.

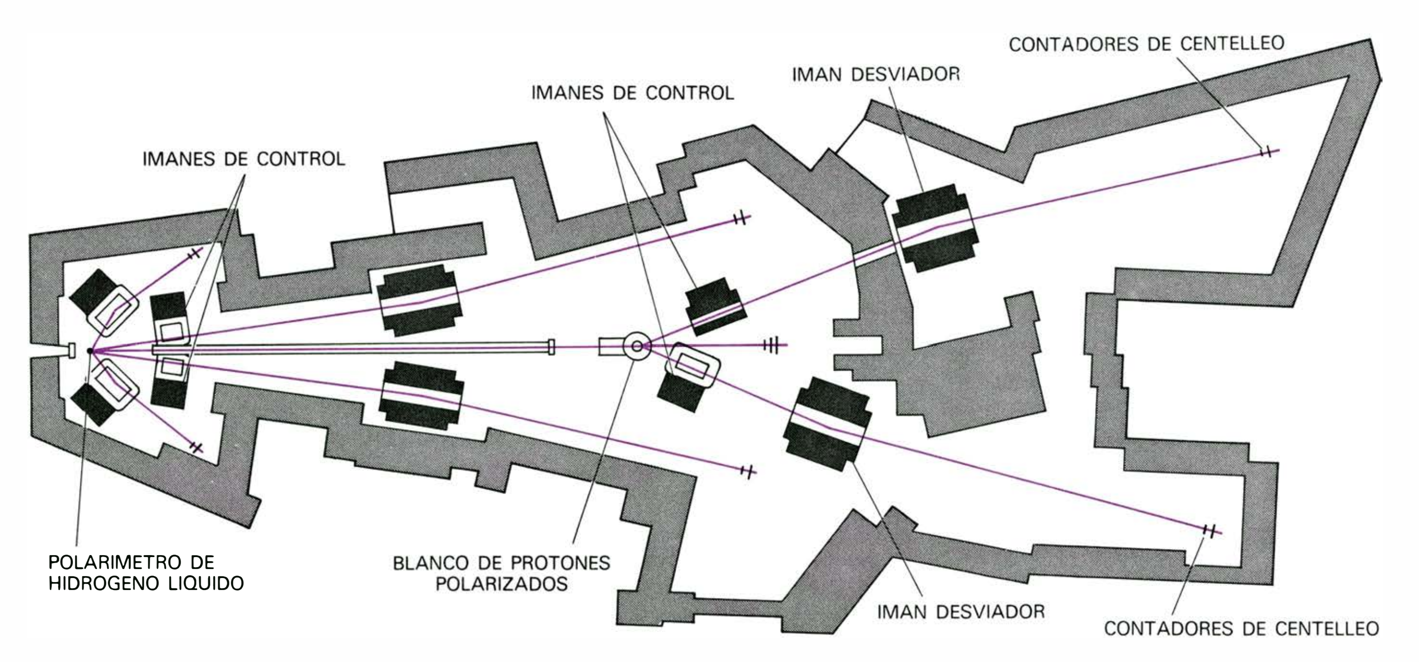
Los spines de los protones están ahora polarizados y los electrones despolarizados. En las etapas finales del procedimiento se arrancan los electrones y se aceleran los protones hasta una energía aproximada de 20.000 electronvolt. La fuente ANAC emite unos 10^{11} protones por segundo con una polarización aproximada de 0,75. La polarización se define como la diferencia entre el número de partículas con spin arriba y el número de partículas con spin abajo, dividida por el número total de protones. Por tanto una polarización de 0,75 implica que la razón de los protones con spin arriba y con spin abajo es de 7 a 1.

La aceleración posterior de los protones tiene lugar en tres etapas. La propia fuente está montada dentro de una cúpula que se mantiene a un potencial de 750.000 volt: cuando los protones dejan la cúpula adquieren por tanto una energía de 750.000 electronvolt. El paso siguiente lo constituye un acelerador lineal de 61 metros de largo, donde los protones son llevados a una energía de 50 MeV (millones de electronvolt). Finalmente se inyectan en el anillo del sincrotrón de protones principal.

El acelerador de Argonne se llama sincrotrón de gradiente cero porque los campos magnéticos que confinan las partículas son campos totalmente uniformes; el ZGS no tiene los imanes cuadrupolares y sextupolares que se emplean en muchos otros aceleradores para la focalización. Como en todos los otros sincrotrones, la aceleración efectiva se realiza sólo en una cavidad comparativamente pequeña, donde un campo de radiofrecuencia establece una diferencia



MECANISMO DE POLARIZACION DEL BLANCO. Es indirecto, como el de la fuente del haz: primero se alinean los spines de los electrones y después la polarización se transfiere a los protones. Se adopta este método porque el electrón tiene un momento magnético unas 660 veces mayor que el del protón y así se polariza más fácilmente. Estos electrones, desapareados de átomos de cromo, se alinean en la configuración de spin arriba mediante un campo magnético fuerte que se aplica al blanco. Los electrones del cromo y los protones de los átomos de hidrógeno se aparean temporalmente de manera que los spines de los electrones están todos arriba y los spines de los protones están distribuidos al azar. Una transición hiperfina inducida por una radiación de microondas invierte los spines de los electrones y protones, pero sólo en los pares en que los spines son opuestos. Cuando los pares se rompen, los protones quedan en el estado de spin arriba y los electrones quedan despolarizados. El proceso es continuo: los spines de los electrones se realinean y transfieren la polarización a los protones cercanos.



ESPECTROMETRO DE DOBLE BRAZO, para medir el ángulo y la energía del protón dispersado y del de retroceso después de una colisión. El haz incide en un blanco de hidrógeno líquido no polarizado y se controlan los sucesos de colisión en cuanto a una asimetría izquierda-derecha que mide la polarización del haz. En el blanco polarizado, los protones dispersados en determinados ángulos pasan a través de una serie de campos magnéticos y se detectan mediante contadores de centelleo. El experimento mide la probabilidad de que un protón se disperse en cada ángulo comprendido entre cero y 90 grados para cada configuración de spin. En vez de mover los detectores para interceptar las partículas dispersadas en los distintos ángulos, unos imanes de control desvían los protones hasta los detectores inmóviles. Los imanes deflectores miden la energía de los protones.

SECCION EFICAZ DE COLISION — 10^5, 10^4, 10^3, 10^2, 10, 1
ANGULO DE COLISION (GRADOS) — 0, 30, 60, 90
DENSIDAD
DISTANCIA AL CENTRO (FERMI) — 0, .5, 1, 1.5

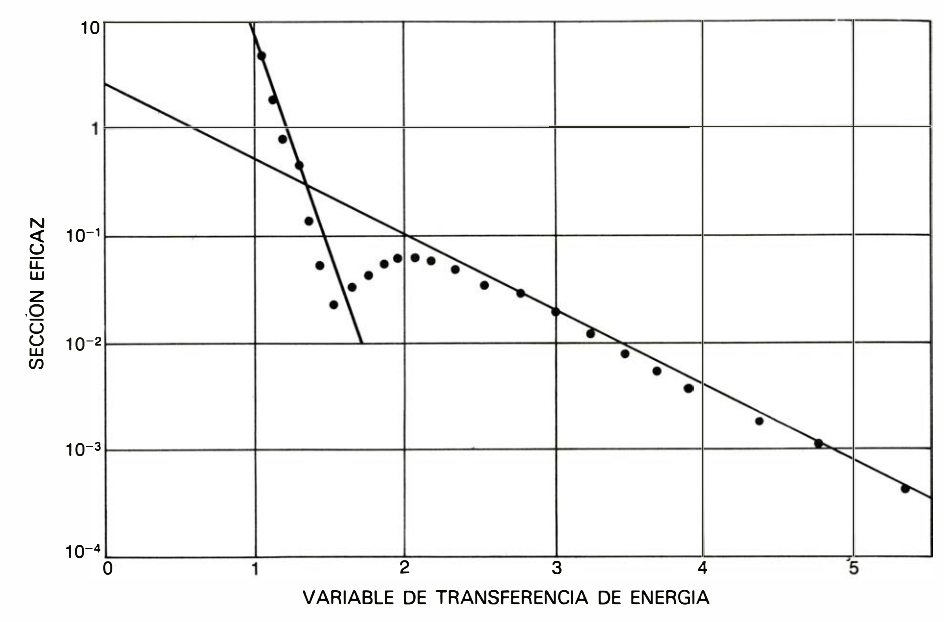
SECCION EFICAZ DE COLISION — 10^5, 10^4, 10^3, 10^2, 10, 1
ANGULO DE COLISION (GRADOS) — 0, 30, 60, 90
DENSIDAD
DISTANCIA AL CENTRO (FERMI) — 0, .5, 1, 1.5

SECCION EFICAZ DE COLISION — 10^5, 10^4, 10^3, 10^2, 10, 1
ANGULO DE COLISION (GRADOS) — 0, 30, 60, 90
DENSIDAD
DISTANCIA AL CENTRO (FERMI) — 0, .5, 1, 1.5
ESTRUCTURA POSIBLE

POSIBLES SECCIONES EFICACES para colisión elástica protón-protón, representadas en función del ángulo de dispersión. La sección eficaz en cualquier ángulo dado mide la probabilidad de que un protón se desvíe dicho ángulo en un choque. Todas las secciones eficaces decrecen rápidamente cuando aumenta el ángulo, lo que simplemente significa que es mucho más probable que un protón sufra una desviación pequeña que no grande. La forma de la curva contiene información acerca de la estructura interna del protón. Un decrecimiento suave de la sección eficaz (*arriba*) sugeriría que la densidad del protón disminuye continuamente desde el centro a la superficie; describiría una partícula "blanda". Una sección eficaz con varios picos y depresiones pronunciados (*en medio*) aparecería si el protón fuera una esfera de contorno bien definido, en la que la densidad es constante desde el centro a la superficie y entonces cae rápidamente a cero. Una sección eficaz con dos componentes separadas por una flexión en la que la pendiente de la curva cambia (*abajo*) correspondería a una estructura con dos regiones distintas: un corazón denso y una nube periférica más difusa. La sección eficaz de colisión elástica del protón se parece a la curva de dos componentes. Del ritmo con que decrece se deduce el tamaño de cada región.

de potencial de unos 12.000 volt. El resto de la máquina, que tiene una circunferencia de casi 170 metros, consiste en imanes que curvan las trayectorias de los protones formando una trayectoria circular. De esta manera cada protón puede pasar muchas veces por la cavidad acelerante, recibiendo un empuje de 12.000 electronvolt de energía cada vez. Para alcanzar 12 GeV un protón debe dar cerca de un millón de vueltas alrededor del ZGS.

El ZGS no fue proyectado para la aceleración de protones polarizados, pero la uniformidad excepcional de sus campos magnéticos simplifica la labor. Los spines de los protones están orientados verticalmente, y cualquier campo horizontal puede despolarizarlos. Incluso en el ZGS hay campos horizontales débiles, pero después de un millón de revoluciones provocan sólo una despolarización de aproximadamente un 1 por ciento. Más serias consecuencias tienen las "resonancias" despolarizantes que se encuentran a determinadas energías en las que la pequeña fuerza despolarizante actúa en la misma dirección en cada vuelta, de manera que se refuerza de forma coherente. Cuando estas fuerzas coherentes se suman en fase, pueden eliminar rápidamente la polarización del haz. Estas resonancias se han evitado instalando dos imanes cuadrupolares y un imán horizontal en el acelerador. Los imanes se conectan bruscamente justo antes de que se alcance cada energía de resonancia y permanecen conectados durante sólo 0,003 segundos. Cambian así ligeramente la trayectoria del haz, permitiendo a los protones saltar rápidamente las energías de resonancia. Estas correcciones programadas con precisión deben aplicarse en 29 resonancias en cada ciclo de aceleración. La técnica de "saltar la resonancia", que se aplicó por primera vez en el ZGS, se ha refinado hasta tal punto que los protones pueden acelerarse hasta 12 GeV manteniendo una polarización de 0,70.



SECCION EFICAZ DE COLISION ELASTICA, medida con protones no polarizados en la Organización Europea para la Investigación Nuclear (CERN) cerca de Ginebra a una energía equivalente a 1500 GeV. Como los spines de las partículas que chocan no se controlaban, la sección eficaz observada es un promedio de todas las secciones eficaces de spin puro. Los datos se representan en función de una variable de transferencia de energía, que mide la violencia de la colisión y está relacionada con el ángulo de desviación. La curva puede descomponerse claramente en dos componentes con pendientes distintas. La porción con pendiente fuerte refleja colisiones leves que afectan sólo a la capa exterior difusa del protón, mientras que la pendiente más gradual del segmento de abajo indica colisiones violentas entre objetos pequeños y duros de dentro del protón. Los objetos parecen tener un radio de alrededor de un tercio de fermi (un fermi es igual a 10^{-13} centímetros). La depresión entre los dos segmentos de la curva puede representar una interferencia entre los dos modos de colisión elástica.

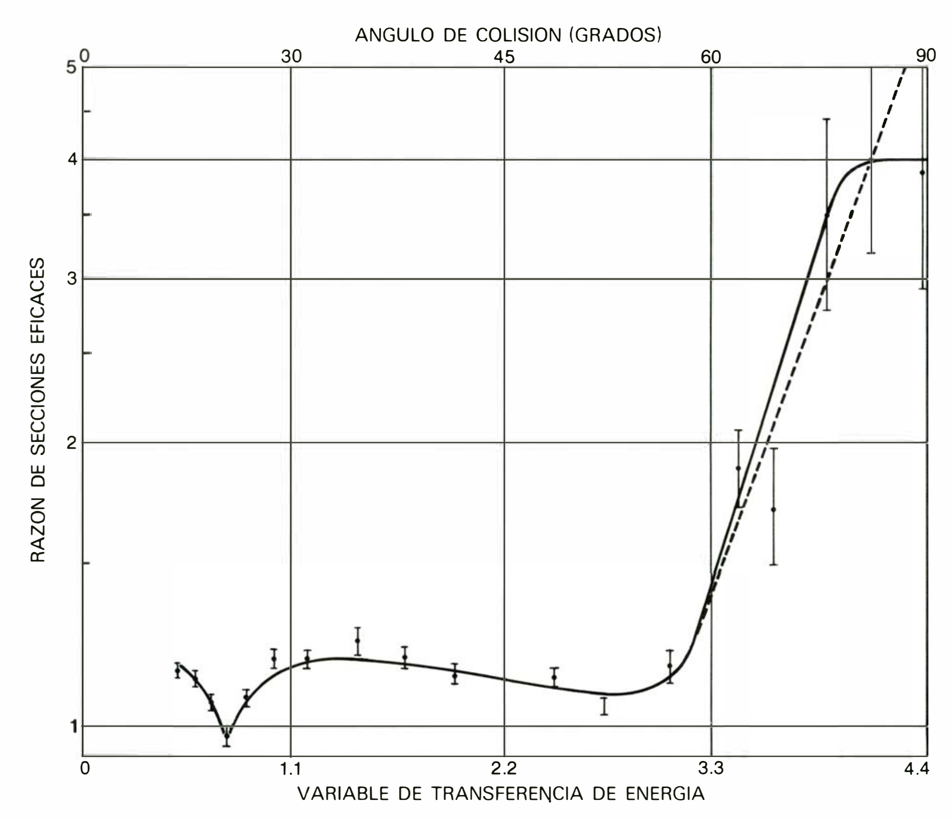
El blanco polarizado funciona de una manera muy parecida a como lo hace la fuente del haz polarizado. La polarización se inicia con un campo magnético fuerte que actúa sobre los electrones, que se polarizan fácilmente debido a su gran momento magnético. Este método de polarización del spin fue desarrollado en los últimos años cincuenta por dos físicos de bajas temperaturas, Anatole Abragam, del Collège de France, y Carson D. Jeffries, de la Universidad de California en Berkeley. Se aplicó en experimentos de física de partículas en Berkeley, en la Organización Europea para la Investigación Nuclear (CERN) y en Argonne. Owen Chamberlain, de Berkeley, y Michel Borghini, del CERN, hicieron importantes contribuciones para adaptar esta técnica a la física de partículas.

La polarización de los protones del blanco es directamente proporcional al campo magnético aplicado y al momento magnético del protón; es inversamente proporcional a la temperatura absoluta. En nuestro blanco el campo aplicado es de 25.000 gauss y la temperatura es de 0,5 grados Kelvin (0,5 grados Celsius por encima del cero absoluto). Bajo estas condiciones los protones aislados tendrían una polarización de sólo 0,003; los electrones, en cambio, la tendrían de 0,995.

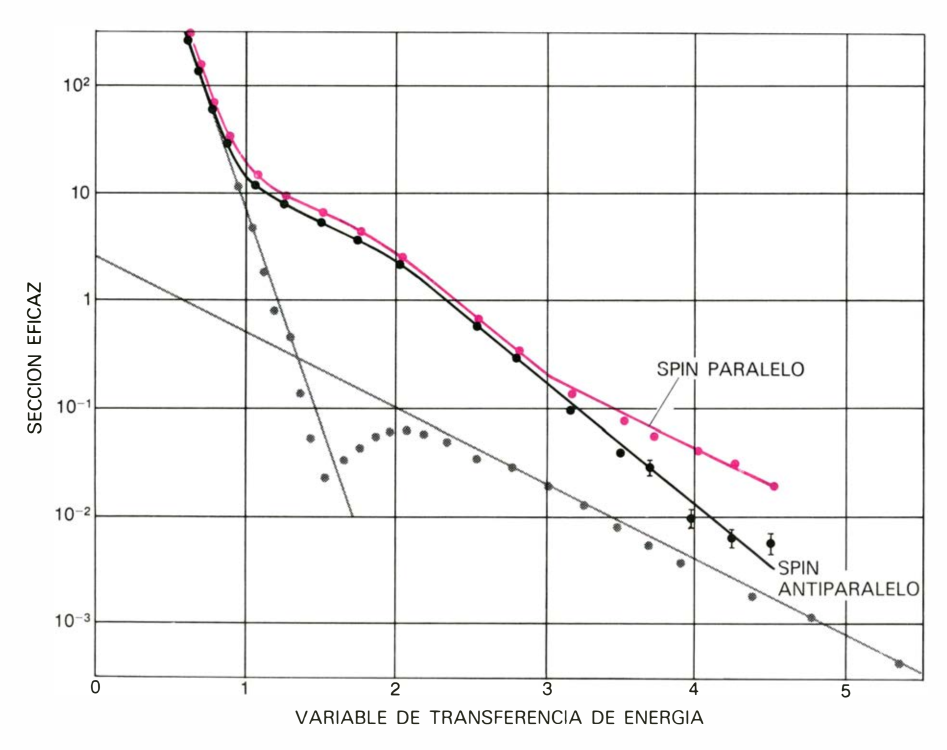
Los protones que se deben polarizar son los núcleos de átomos de hidrógeno de una sustancia orgánica tal como el etilenglicol dopado con cromo. (Irónicamente el etilenglicol es un anticongelante usual, pero en nuestro aparato está congelado formando pequeños granos.) Lo que de hecho polariza el campo magnético aplicado son los electrones desapareados de los átomos de cromo, que pueden aparearse temporalmente con los protones del hidrógeno. En cada par cromo-hidrógeno hay sólo dos estados de spin posibles: electrón y protón con spin arriba, y electrón arriba y protón abajo. Un campo de microondas con una frecuencia de 70 gigahertz invierte entonces ambos spines en los pares del último tipo, despolarizando los electrones y polarizando todos los protones en el estado de spin arriba. La anchura de banda de la señal de microondas debe ser lo suficientemente estrecha para evitar la frecuencia de 70,2 gigahertz, que estimula la transición de inversión doble equivalente en los otros pares de spin.

El enfriamiento lo proporciona un isótopo poco abundante del helio, el helio 3, que circula entre los gránulos de etilenglicol. La temperatura del helio 3 se mantiene a 0,5 grados Kelvin mediante una gran bomba de vacío que reduce la presión de vapor del helio acelerando por tanto su evaporación. El conjunto del blanco, con su refrigerante de helio 3, está inmerso en una vasija de helio 4, el isótopo ordinario, que se enfría a dos grados Kelvin mediante métodos similares.

La polarización del blanco de Argonne ha llegado a ser de 0,85, pero este valor no puede mantenerse cuando el haz está funcionando debido a que la radiación perjudica los gránulos de etilenglicol. Puede lograrse una polarización media de cerca 0,7 templando los gránulos a una temperatura de 160 grados Kelvin cada ocho horas y renovándolos cada varios días.



RESULTADOS DEL ESTUDIO DE MICHIGAN-ARGONNE sobre colisiones de protones polarizados presentados en función de un cociente: la sección eficaz medida cuando los spines de los protones del haz y del blanco eran paralelos dividida por la sección eficaz cuando eran antiparalelos. Es de destacar que esta razón es mayor que la unidad para todos los ángulos de dispersión o colisión con una posible excepción; por tanto es siempre más probable que los protones interaccionen cuando están girando en la misma dirección. El aumento brusco de la razón para un ángulo de unos 60 grados indica que el spin es más importante para las colisiones violentas, que presumiblemente involucran a los objetos duros del interior del protón. De los datos no se deduce claramente si la razón de las secciones eficaces continúa aumentando para ángulos de colisión mayores o se estabiliza en el valor 4.



LOS EFECTOS DEL SPIN en las colisiones protón-protón se aprecian también cuando las secciones eficaces para sucesos de spin paralelo (*color*) y para sucesos de spin antiparalelo (*negro*) se representan por separado. Las dos secciones eficaces se diferencian muy poco para ángulos pequeños, colisiones leves, pero se diferencian en las colisiones de gran ángulo. Los datos son compatibles con la hipótesis de que no hay ninguna flexión en la sección eficaz de colisión de spin antiparalelo, mientras que la de spin paralelo presenta un persistente cambio de pendiente. Para valores grandes de la variable de transferencia de energía la pendiente de la curva de spin paralelo es la misma que en las medidas de 1500 GeV (*gris*).

Una componente esencial de cualquier experimento de colisión, se emplee o no partículas polarizadas, es un espectrómetro: un sistema de detectores y de imanes deflectores (desviadores) dispuestos para medir la posición y la energía de las partículas dispersadas. Nosotros empleamos un espectrómetro de doble brazo: con detectores tanto para los protones dispersados como para los de retroceso. Consiste en cuatro imanes que curvan las trayectorias de los protones y por tanto ayudan a determinar su energía, y seis contadores de centelleo, que detectan la posición angular de una partícula de la misma manera que lo hacía la pantalla fluorescente de Rutherford.

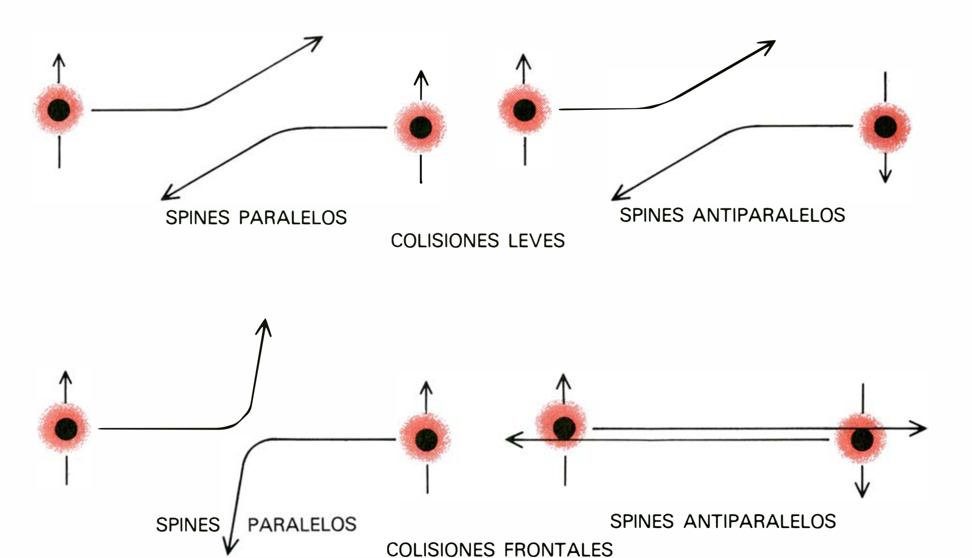
La máxima resolución la proporciona el brazo que intercepta la partícula dispersada. El ángulo de desviación se mide con una precisión de unos 0,5 grados y la energía del protón dispersado se determina con una precisión de un 7 por ciento. La resolución de los otros contadores no es tan grande, pero hacen la importante función de identificar los sucesos extraños de manera que se les pueda rechazar. Estos sucesos de fondo incluyen sucesos accidentales o coincidentes, en los que las partículas que se detectan provienen de dos choques simultáneos e independientes. Otras señales de fondo provienen de colisiones inelásticas y de la dispersión por núcleos de carbono y oxígeno del etilenglicol.

El procedimiento corriente en un experimento como éste consiste en medir una sección eficaz a un ángulo dado y mover entonces los brazos del espectrómetro a lo largo de un arco centrado en el blanco hasta la posición angular siguiente. El movimiento del aparato es especialmente incómodo y largo de realizar en nuestro experimento. (Debe quitarse el techo del edificio.) Simplificamos los cambios de ángulo colocando un imán deflector (desviador) con una gran apertura cerca del blanco en cada brazo del espectrómetro. El imán desviador captura las partículas dispersadas y las guía hasta los contadores de centelleo fijos. Puede observarse así un gran intervalo de ángulos sin mover ninguna parte del aparato; sólo debe cambiarse la corriente del imán.

Contamos sucesos en ángulos de colisión comprendidos entre cero y 90 grados, medidos con respecto a la dirección del haz incidente en el sistema de coordenadas del centro de masas. (Los ángulos observados en el sistema de referencia de laboratorio son más pequeños.) Los protones pueden desviarse a ángulos mayores que 90 grados; de hecho las colisiones más violentas son las de 180

grados, cuando el protón dispersado rebota directamente hacia atrás a lo largo del eje del haz. Como todos los protones son idénticos, sin embargo, no es posible determinar cuál es el protón dispersado y cuál es el protón de retroceso. A fin de evitar esta ambigüedad clasificamos los sucesos de colisión según el ángulo del protón que sale hacia adelante, que debe ser siempre de 90 grados o menos.

El experimento se hizo durante las 24 horas del día a lo largo de unos seis meses, detectando uno o dos sucesos por hora en ángulos de colisión grandes. El spin de los protones del haz se invertía con cada ciclo del sincrotrón (aproximadamente una vez cada tres segundos), y la polarización del blanco se invertía cada vez que los gránulos eran templados o sustituidos (aproximadamente una vez cada ocho horas). También se hicieron pruebas con haces no polarizados, con blancos no polarizados y con blancos sin hidrógeno. La consistencia de los resultados obtenidos bajo estas distintas condiciones asegura que no había errores sistemáticos en el aparato. Las polarizaciones del haz y del blanco también se controlaban continuamente. Debido a que las polarizaciones eran siempre menores que la unidad corregimos los datos recogidos para deducir las secciones eficaces de spin puro.



INTERPRETACION DE LOS EFECTOS DE SPIN observados en experimentos de colisiones elásticas. Dicha interpretación sugiere que la mayor parte del momento angular de spin del protón se debe a los objetos interiores de un tercio de fermi. Cuando interaccionan las partes más externas del protón, como sucede en colisiones leves (*arriba*), la influencia del spin es pequeña y las dos secciones eficaces son casi iguales. En colisiones frontales, en las que participan los objetos duros (*abajo*), el spin es de importancia decisiva. Cuando los spines de los dos protones son paralelos, los objetos de un tercio de fermi rebotan violentamente; cuando los spines son antiparalelos, los objetos parecen pasar de largo. Se muestra el protón como si tuviera un único corazón denso, una estructura propuesta por el autor en 1963. Los datos no excluyen la posibilidad de que haya más de uno de estos objetos en cada protón.

Como he indicado antes, una sección eficaz medida a un ángulo definido, representa la probabilidad de que dos protones que chocan se dispersen elásticamente según dicho ángulo. La variación de la sección eficaz cuando el ángulo cambia proporciona información acerca del tamaño, forma y estructura interna del protón. Sólo las colisiones violentas que dan lugar a dispersión en ángulos grandes pueden informar sobre los detalles más pequeños de la estructura del protón. Son los sucesos de colisión a gran ángulo los que pueden revelar la presencia de cualquier objeto duro dentro del protón.

Todos los datos recogidos en una experiencia de colisiones elásticas con protones no polarizados pueden reducirse a un único gráfico, en el que la sección eficaz se representa en un eje y el ángulo de colisión en el otro. Cuando se miden las secciones eficaces de spin puro, se debe dibujar un gráfico distinto para cada estado de spin. La información acerca de la estructura del protón se extrae de estos gráficos por el método matemático llamado análisis de Fourier. No daremos aquí los detalles de esta técnica, pero algunas de las conclusiones que se sacan de ella son bastante intuitivas.

Una forma que puede tener concebiblemente el gráfico de la sección eficaz es la de una línea recta que desciende desde una sección eficaz grande a cero grados hasta un valor mínimo a 90 grados. Un tal gráfico sin estructura sugeriría que el protón es una esfera blanda y suave sin bordes abruptos y sin objetos duros en su interior. La densidad bajaría suavemente desde un valor máximo en el centro hasta cero en un radio de un fermi aproximadamente.

Otra posibilidad sería una curva con la misma característica global pero con picos y depresiones pronunciados superpuestos sobre ella. Un tal gráfico sería análogo a la figura de difracción formada cuando un haz de luz es interceptado por un objeto de borde bien definido; el análisis de Fourier del gráfico describe al protón como una esfera de bordes bien definidos. La densidad es esencialmente constante desde el centro hasta un borde cercano al radio de un fermi; allí baja rápidamente hasta cero.

Un tercer gráfico posible se describe mejor como una mezcla de dos curvas distintas. Para pequeños ángulos de colisión (digamos menores de 60 grados) la sección eficaz decae suave y rápidamente. En 60 grados hay una flexión en la curva, y para ángulos mayores la sección eficaz decrece más lentamente. Una tal curva de dos componentes sugiere que el protón tiene dos regiones, la más interna densa y la externa difusa; la diferencia entre las dos curvas indica que las dos regiones están claramente definidas. A partir del ritmo con que decrece la sección eficaz podemos calcular el tamaño de cada región. El gráfico real de la sección eficaz para dispersión elástica protón-protón tiene una forma muy parecida a este tercer ejemplo, con dos curvas separadas por un cambio brusco en la pendiente.

En el año 1963 sugerí que el protón tiene un "corazón" rodeado por una "nube" de menor densidad. De hecho podían haber varias capas de densidad decreciente, por lo que se conoce mi propuesta como modelo de la cebolla. La primera prueba experimental de la estructura interna del protón fue obtenida en 1966 en un experimento con protones no polarizados realizado en el ZGS por mí y mis colegas de Michigan y Argonne. El experimento tenía una forma ligeramente distinta de los descritos antes. En vez de estudiar sucesos de una única energía de haz y varios ángulos de colisión, manteníamos el ángulo constante a 90 grados y medíamos la sección eficaz cuando la energía de la partícula incidente variaba de 5 a 13 GeV. A energías de hasta 8 GeV la sección eficaz decrecía rápida y suavemente; en dicha energía, sin embargo, se producía un cambio brusco pasando a una pendiente más gradual. Esta brusca flexión en la curva de la sección eficaz se interpretó como prueba de la existencia de objetos dentro del protón de un tamaño de un tercio de fermi.

Experimentos recientes con los "anillos de almacenamiento" (Intersecting Storage Rings) del CERN confirman la

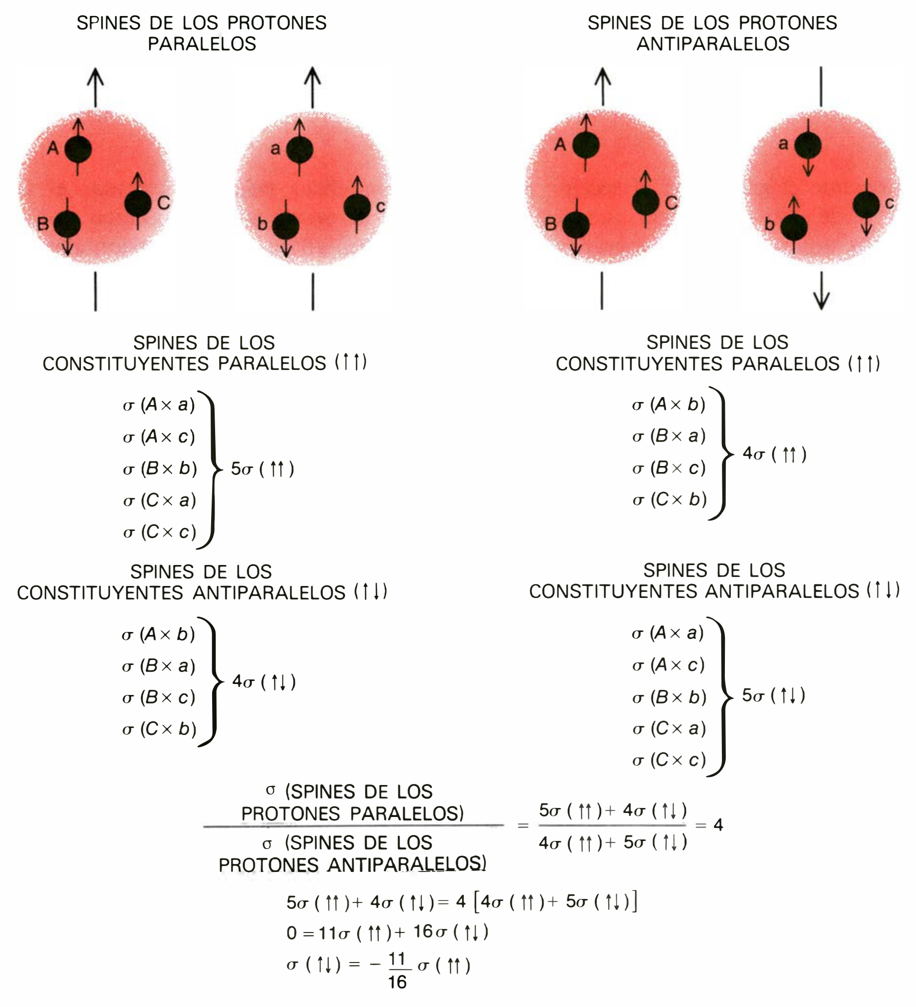
hipótesis de estos objetos de un tercio de fermi. En los anillos de almacenamiento los protones circulan en direcciones opuestas, de manera que chocan frontalmente. De esta manera se logran energías de centro de masas extraordinariamente elevadas; para alcanzar resultados equivalentes en un acelerador de blanco fijo se precisaría una energía de haz de 1500 GeV. Los investigadores del CERN observaron una flexión distinta en la curva de la sección eficaz cuando pasaban de colisiones de pequeño ángulo a colisiones de gran ángulo. Realmente, a estas energías altas la distinción entre el corazón y la nube parece ser bastante definida. Puede ser particularmente significativo que el tamaño aparente del corazón es el mismo en los análisis a 1500 GeV que el encontrado en el experimento entre 5 y 13 GeV.

En 1970 surgieron otras pruebas de la estructura interna a raíz de las experiencias del Stanford Linear Acceleraton Center (SLAC) donde se hacían chocar electrones de 20 GeV contra un blanco de protones en reposo. Desde entonces han finalizado en varios sitios experimentos similares con neutrinos de alta energía dispersados por protones. En estos experimentos se contaron todas las interacciones, no sólo las elásticas. Tales experimentos se llaman "inclusivos"; aunque las secciones eficaces resultantes son algo más difíciles de interpretar, a menudo son mucho más fáciles de medir que la sección eficaz elástica. En SLAC se detectó un número sorprendente de sucesos inclusivos de gran ángulo, lo que sugiere que el protón tiene una estructura interna de algún tipo, pero no hay una manera directa de calcular el tamaño de estos objetos a partir de los datos de las colisiones.

A pesar de las dificultades de interpretación, los experimentos de dispersión de electrones y de dispersión de neutrinos se han aceptado ampliamente como prueba de ciertos modelos de estructura del protón. Uno de estos modelos, introducido en 1968 por Richard P. Feynman, del Instituto de Tecnología de California (Cal Tech), propone que el protón está formado por muchos de los diminutos objetos llamados partones. Otro modelo es la teoría de quarks, formulada en 1963 por Murray Gell-Mann y George Zweig, que también estaban en el Cal Tech. En este modelo se supone que el protón, el neutrón y muchas partículas relacionadas representan cada una distintas combinaciones de tres quarks. El modelo ha tenido bastante éxito a la hora de explicar la clasificación de varias grandes familias de partículas y ha servido de guía en la búsqueda de nuevas partículas.

A partir de los resultados de experimentos de colisión durante 15 años creo que está razonablemente bien establecido que el protón tiene algunos objetos constituyentes más pequeños. Los experimentos de colisión elástica sugieren fuertemente que el tamaño de estos objetos es de alrededor de un tercio de fermi. En mi opinión, todas las otras propiedades de estos objetos son muy inciertas, incluyendo su forma, masa, spin y carga eléctrica. Hasta el mismo número de objetos constituyentes se mueve en el terreno de conjetura. No se advierten pruebas directas que relacione los quarks, que han sido útiles para explicar las familias de partículas, con los objetos que se ven en los experimentos de colisión. Debo puntualizar, sin embargo, que me encuentro entre la minoría que mantiene esta opinión.

Los experimentos en que se miden las secciones eficaces de spin puro proporcionan una nueva prueba para los modelos de la estructura del protón. En los dos últimos años hemos medido secciones eficaces de spin puro a una energía de haz incidente de 12 GeV y a ángulos de colisión (en el sistema de coordenadas de centro de masas) variando desde cerca de cero hasta exactamente 90 grados. Conociendo la polarización del haz y de los protones del blanco, pero sin determinar los spines de las partículas dispersada y de retroceso, hemos podido distinguir cuatro estados de spin: haz con spin arriba y blanco con spin abajo, haz con spin abajo y blanco con spin arriba, ambos spines arriba y ambos spines abajo.



LA HIPOTESIS DE LOS QUARKS sugiere otro modelo para la estructura del protón, pero éste no puede reconciliarse fácilmente con los resultados de los experimentos de colisión de spin determinado. Si los quarks son los objetos de un tercio de fermi observados, entonces la sección eficaz para ángulos grandes en colisiones de protones debe igualarse a la suma de las secciones eficaces para todos los sucesos en los que un quark es dispersado por otro. Se supone que cada protón consta de tres quarks, de manera que hay nueve secciones eficaces para colisión quark-quark. Se designa aquí a la sección eficaz por la letra griega sigma (σ), y los símbolos ↑↑ y ↑↓ representan respectivamente sucesos de spin paralelo y sucesos de spin antiparalelo. Como la sección eficaz del protón para las dos configuraciones de spin están en una razón de 4 a 1, las sumas de todas las secciones eficaces de quarks deben tener la misma razón. Si se hace la razón para las colisiones quark-quark igual a 4, se llega a un absurdo: una de las secciones eficaces de quarks debe ser negativa, y una probabilidad no puede ser menor que cero. La dificultad puede evitarse sólo en modelos con interacciones quark-quark más complejas.

Las secciones eficaces para los dos primeros de estos estados en que los spines son antiparalelos deben ser, por principios de simetría, exactamente iguales. El experimento confirma esta predicción, proporcionando por tanto otra prueba para los errores sistemáticos. Las secciones eficaces correspondientes al caso de los dos spines paralelos, ambos hacia arriba o ambos hacia abajo, no tienen que ser necesariamente iguales. Encontramos alguna diferencia entre ellas en la colisión a ángulos pequeños y medianos, pero en las colisiones más violentas son casi exactamente iguales.

Nuestros hallazgos más importantes se representan adecuadamente agrupando todos los datos medidos en dos secciones eficaces, una para todos los estados en los que las partículas entrantes tienen spines paralelos, y una para los estados correspondientes con spines antiparalelos. Estas dos cantidades son promedios de cuatro secciones eficaces de spin puro, pero parecen mostrar los mayores efectos de spin observados hasta ahora en colisiones protón-protón.

Los resultados de nuestro experimento de dos años pueden recogerse sucintamente en un gráfico que represente la razón de la sección eficaz de spin paralelo a la sección eficaz de spin antiparalelo en función del ángulo de dispersión o colisión. Incluso para ángulos medios o pequeños (hasta 60 grados), la curva adquiere rasgos interesantes. El primero es simplemente que la razón no es nunca menor que la unidad: independientemente de cuál sea el ángulo de colisión, la colisión de spin paralelo es más probable que la colisión de spin antiparalelo. Por otro lado, para ángulos de menos de 60 grados, la sección eficaz paralela no es nunca más del 25 por ciento mayor que la sección eficaz antiparalela. El hecho más curioso en esta región es una estrecha depresión a unos 25 grados: sólo para este ángulo las dos probabilidades de colisión son iguales. La depresión es tan estrecha y tan manifiesta que parece que debe tener alguna explicación simple, pero, por ahora, esta explicación simple no existe.

Para ángulos pequeños, la forma detallada de la curva es intrigante, pero la magnitud aproximada de los efectos del spin no es excepcional. Para mayores ángulos de dispersión, sin embargo, sorprende el comportamiento de la razón. Los sucesos de gran ángulo de dispersión son choques violentos, en los que las partículas intercambian una gran cantidad de energía. Habíamos esperado que la pequeña energía del spin quedaría enmascarada en estos sucesos por la enorme energía de la colisión y que el spin no sería importante. En cambio, la influencia del spin aumenta dramáticamente. Para ángulos mayores de 60 grados, la razón de la sección eficaz paralela a la antiparalela se dispara hacia arriba y para 90 grados alcanza un valor de 4. Puede subir todavía más: la incertidumbre en las medidas es aún demasiado grande para que podamos determinar si el 4 es o no el valor límite.

Los mismos datos pueden presentarse en la forma de curvas separadas para cada sección eficaz. Algunas de las propiedades observadas en los anteriores experimentos de 1500 GeV con protones no polarizados se ponen entonces de manifiesto. Lo más notable es que la curva de spin paralelo presenta la forma de dos componentes familiar, con una aguda flexión que separa los dos segmentos. Además, la pendiente del segmento de ángulo grande, que corresponde a sucesos de colisión fuerte, da el mismo resultado para el tamaño de los objetos interiores: un tercio de fermi.

Hay otro hecho remarcable de estas dos curvas. La transición brusca en la sección eficaz de spin paralelo sucede a un ángulo de colisión de unos 60 grados: el mismo ángulo en el que la razón de la sección eficaz paralela a antiparalela sube repentinamente. Aunque los datos no son aún bastante precisos para la certeza, puede ser que no exista en absoluto la flexión correspondiente en la sección eficaz de spin antiparalelo. Si es así, el aumento repentino de la razón podría explicarse fácilmente. A un ángulo de dispersión de unos 60 grados, los objetos de un tercio de fermi se hacen de la máxima influencia en la colisión elástica protón-protón, y estos objetos aparentemente chocan con violencia sólo en los protones de spin paralelo.

Si estos resultados preliminares resultan ser correctos, la interpretación de las secciones eficaces de spin determinado es simple, pero sorprendente. Cuando dos protones girando en la misma dirección chocan frontalmente, casi siempre se dispersan de un modo violento. Cuando las mismas partículas están girando en direcciones opuestas, a menudo pasan de largo como si fuesen transparentes. Las interacciones a estos ángulos grandes se deben probablemente, casi por completo, a los objetos de un tercio de fermi. Presumiblemente estos objetos están girando muy deprisa y por alguna razón rara vez interaccionan a menos que sus spines sean paralelos.

La magnitud del efecto del spin puede revelar algo acerca de la naturaleza de los objetos interactuantes. Para el ángulo de dispersión máximo (90 grados) la contribución de los objetos duros parece ser máxima. Para dicho ángulo, la razón de las secciones eficaces paralela a antiparalela resultó ser 4, y este valor puede reflejar de alguna manera el número de objetos en el protón y la naturaleza de sus spines.

Para el modelo de cebolla, con un único corazón duro, estas observaciones simplemente sugieren que el corazón lleva la mayor parte del momento angular de spin del protón. Las interacciones que involucran sólo la nube exterior sólo están influenciadas ligeramente por el spin. El valor numérico de 4 observado en el experimento se explica como la razón de las secciones eficaces de spin paralelo a spin antiparalelo para interacciones directas entre los corazones. Para ángulos de dispersión menores, la colisión directa corazón-corazón queda obscurecida por la influencia moderadora de la nube que gira lentamente.

El análisis es más complicado, y aparentemente menos satisfactorio, para un protón formado por tres quarks. Cada quark debe tener un spin de $1/2\ \hbar$; como el protón en conjunto tiene el mismo valor del momento angular de spin, los quarks deben siempre combinarse con dos de sus spines paralelos y un spin en la dirección opuesta. Las colisiones de gran ángulo se deben presumiblemente a la colisión directa de un quark del protón del haz con un quark del protón del blanco. Para tales colisiones quark-quark las secciones eficaces de spin puro para dispersión protón-protón deben ser iguales a la suma de las secciones eficaces de spin puro para las nueve combinaciones posibles de los quarks que chocan. Cuando se calcula la razón de las dos sumas, debería dar el valor 4 observado en las colisiones protón-protón. Esta evaluación conduce a un absurdo: una sección eficaz negativa, o en otras palabras una probabilidad de menos de cero. Un modelo más complejo para la colisión quark-quark puede eliminar esta dificultad, pero también comprometería la elegancia del propio modelo.

La gran e inesperada influencia del spin en las colisiones con gran ángulo de dispersión, limita fuertemente cualquier teoría que explique los experimentos de colisiones violentas en términos de los constituyentes fundamentales del protón. Mis colegas y yo esperamos extender pronto los estudios a colisiones de mayor violencia, en las que la estructura interior del protón pueda manifestarse aún más claramente.

La conquista de Europa occidental por las lenguas indoeuropeas

Las grandes lenguas o familias de lenguas de Occidente, así el celta o el germánico, se extienden relativamente tarde hasta las orillas del Atlántico. El vasco es el único sobreviente del mundo lingüístico anterior

Antonio Tovar

Para mis investigaciones de lingüística comparada tuvo una significación decisiva el trabajo de celtista J. Pokorny. En un artículo del año 1940, publicado en Irlanda y dedicado a interpretar el nombre de la isla, se empezaba diciendo que Hispania, la Península Ibérica, era una clave para rastrear los pueblos primitivos de la más remota Europa occidental. Y esto por dos razones: la primera, que en sus exploraciones los pueblos de alta cultura de la antigüedad recorrieron antes las costas de Hispania que las de los países atlánticos, y por eso las noticias que podemos leer en las fuentes griegas y romanas son más antiguas y copiosas; la importancia geográfica del estrecho de Gibraltar y las riquezas minerales de la Península atrajeron a pueblos navegantes prehistóricos, y la historia nos habla de la presencia muy temprana de fenicios y griegos. La segunda razón es que, como la Península es el extremo sudoccidental, las migraciones procedentes del centro y norte de Europa debieron llegar menos numerosas y con menos superposiciones, es decir, cabe esperar que más fáciles de analizar en sus restos arqueológicos y lingüísticos.

A estas consideraciones que dedujimos de la indicación de Pokorny hemos de añadir otra: a ambos lados de los Pirineos, en la zona comprendida entre las ciudades de Bilbao, Bayona y Pamplona, más de medio millón de hablantes han conservado la antigua lengua del país, el euskera o vascuence, que no es indoeuropea, sino preindoeuropea: único resto que queda como testigo de lo que fue lingüísticamente la Europa occidental anterior, no sólo a los romanos, que llegaron hace dos mil años, sino a los celtas y otros supuestos pueblos afines, que comenzaron la indoeuropeización del Occidente mil años antes.

El estudio, pues, de las antigüedades lingüísticas de la Península, en el que se ha progresado bastante en el último medio siglo, permite intentar una interpretación de lo que se sabe de la formación de las lenguas indoeuropeas y sobre todo de su expansión hacia occidente.

Las lenguas indoeuropeas

Es sabido que casi todas las lenguas de Europa y algunas de Asia occidental y meridional pertenecen a una misma familia, conocida generalmente con el nombre de indoeuropea. Son hermanas o, si seguimos la metáfora, primas o parientes entre sí, todas las lenguas latinas, las germánicas, las eslavas, el griego y otras menores, como las celtas que sobreviven en Irlanda, Gales y Bretaña, el albanés y el lituano, el armenio, y por extensos territorios de Asia y con millones de hablantes tenemos el persa o iranio, con otros dialectos, y las numerosas lenguas de la India que se agrupan como arias. Pertenecieron también a esta familia muchas lenguas que se extinguieron, dejando o no restos escritos, alguna de ellas, como el hetita en Asia menor, importantísima por la antigüedad de sus textos, que comienzan en el siglo XVII a. de C.

Si además pensamos en la gran difusión actual de algunas de estas lenguas, como el inglés, el español, el ruso, el portugués, el francés, podremos decir, sin pecar de europeocentrismo, que es el grupo lingüístico más importante del planeta.

La semejanza de todas estas lenguas se explica por un origen común. Desde 1816 el parecido empezó a estudiarse metódicamente, cuando F. Bopp demostró la semejanza de la conjugación del sánscrito con la del griego, latín, persa y germánico, y aproximadamente al mismo tiempo R. Rask señalaba el parentesco de algunas de estas lenguas con las bálticas y eslavas.

A lo largo de siglo y medio los cultivadores de la lingüística histórica comparada han reunido una cantidad impresionante de hechos que, debidamente sistematizados, permiten reconocer el origen común de todas esas lenguas y los estadios de su diferenciación ulterior.

La "reconstrucción", el método desarrollado por los lingüistas de dirección histórica, permite llegar en cierta medida a la lengua común, es decir, a tiempos anteriores al uso de la escritura. Así como comparando el español, portugués e italiano *padre,* francés *père,* catalán *pare,* etc. se reconstruye el acusativo latino *patre(m)* como forma originaria, comparando el latín *pater* con el griego *patér,* sánscrito *pitá,* irlandés *athir,* se puede construir una forma **pətēr* que se escribe con un asterisco para indicar que no la poseemos así en ningún texto y se supone es la que se usó en los remotos tiempos en que los indoeuropeos no se habían separado aún. Esas formas con asterisco, puramente reconstruidas sobre las de distintas lenguas conocidas más tarde, se supone se oyeron de labios de gentes que vivieron hace más de 5000 años.

Se comprende que los comparatistas, ocupados en el trabajo de estudiar las diferentes lenguas y de seguir su evolución, se limitaran en general por de pronto a dejar en un abstracto limbo el problema de la cronología y de la ubicación geográfica de los antiguos indoeuropeos.

El problema de la diferenciación de las lenguas se explicó primero sin consideraciones de tiempo ni espacio con los conocidos modelos de la división en "ramas", como un árbol genealógico (A. Schleicher), o, más de acuerdo con la división dialectal de lenguas vivas, de la diferenciación con "ondas" (J. Schmidt), comparándolas a las que producirían diversas piedras (como centros de dife-

renciación e iniciativa política, cultural, etcétera) al caer en diversos puntos de un estanque.

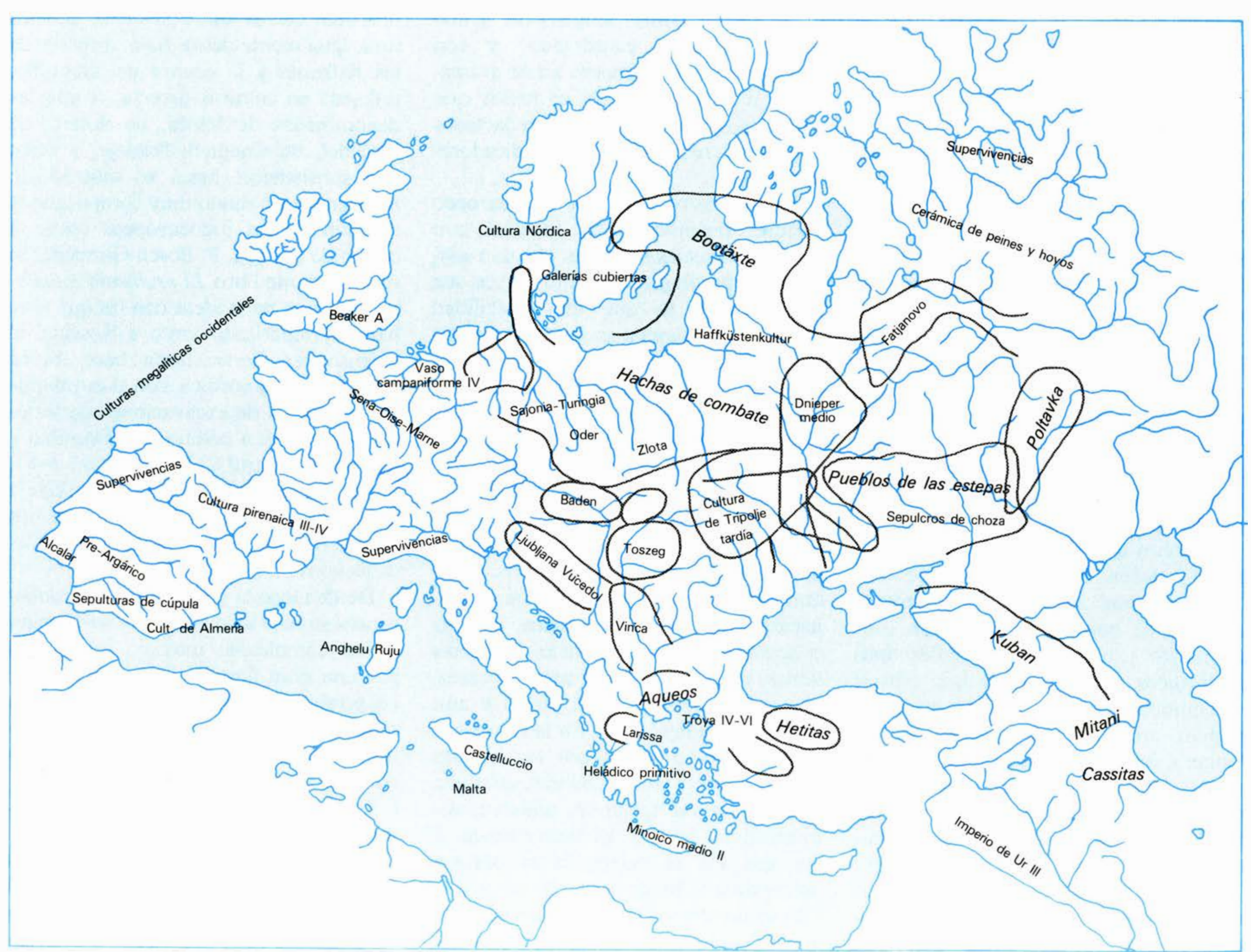
La imagen abstracta de la diferenciación o separación dentro de la familia indoeuropea se fue planteando de modo más concreto al aplicar lo que se fue sabiendo con los estudios dialectológicos en lenguas vivas (románicas, germánicas, etc.), y entonces se hizo más tentadora la comparación de los movimientos de pueblos que podían suponerse con las culturas que identificaba la arqueología prehistórica. Se comprende que muchos lingüistas y muchos prehistoriadores se resistan a acercar ambos campos de estudio, pues mientras los materiales arqueológicos no van acompañados de inscripciones o no corresponden a épocas documentadas por la historia, resulta arriesgada la identificación de pueblos y hablantes de lenguas determinadas. La experiencia enseña que la difusión de una lengua no va siempre acompañada de un cambio visible en la cerámica o los usos funerarios. Los peligros de la identificación son tanto mayores cuanto más nos remontamos en el tiempo y más nos alejamos de los testimonios históricos o lingüísticos. Y sin embargo, aun con riesgo de cometer errores, son muchos los estudiosos (más entre los prehistoriadores que entre los lingüistas) que se atreven a intentar síntesis. Nosotros nos basamos en algunas recientes e intentamos completarlas en lo referente a la indoeuropeización de Europa occidental.

La "patria" de los indoeuropeos

Los comparatistas, basándose puramente en hechos lingüísticos, fijaron, ya desde mediados del pasado siglo, para el foco desde el que se expandieron los indoeuropeos, una serie de peculiaridades climáticas y culturales que permitían situar la patria (*Urheimat*) del supuesto pueblo en un clima templado, más bien frío, en el que, por ejemplo, faltaban leones, tigres y vid; podía suponerse una ganadería muy desarrollada, pero la agricultura era más rudimentaria. Tal época común queda datada porque el único nombre de metal común a varias lenguas en un área muy difundida es "bronce" o "cobre" (latín *aes*).

Como se consiguió antes en el Occidente de Europa un cuadro comprensible, gracias al progreso de las excavaciones, los prehistoriadores alemanes y escandinavos se inclinaron a menudo a poner en el Norte la patria de los indoeuropeos, lo cual combinado con ideas racistas, llevó a la identificación del "ario" (la palabra *arya* en sánscrito significa señor) con el rubio nórdico. Y lo que había comenzado como hipótesis científica apoyada quizás en inconsciente o consciente nacionalismo, terminó en la catástrofe de Hitler. Ante la exageración racista de identificar pueblo con lengua, muchos comparatistas se apresuraron a negar que los "indoeuropeos" de la etapa primitiva común fueran un verdadero pueblo con una verdadera lengua.

Después de la segunda guerra mundial el progreso de las excavaciones y estudios arqueológicos en los países de Europa oriental ha permitido nivelar los conocimientos con los de los países



LOS PRIMEROS PUEBLOS INDOEUROPEOS que aparecen en la historia (hetitas, griegos aqueos, los indoiranios en su migración) se extienden hacia el sur y el sudeste. El mapa refleja consecuencias de su penetración en la antigua Europa. Bosch veía indoeuropeos no sólo en la zona póntico-caucásica, sino más al oeste. Los pueblos de las hachas de combate aparecen como agentes de la indoeuropeización en la Europa nórdica y central.

occidentales, y el difícil problema de acercarse arqueológicamente a la "patria primitiva" de las gentes que hablaron el "indoeuropeo común" puede plantearse con datos nuevos. Los supuestos culturales a que los lingüistas habían llegado: sociedades guerreras, patriarcales, con una división en clases, organizadas en pequeños grupos con jefes poderosos, dueñas de caballos y de grandes rebaños, que habitaban en aldeas y tenían también castillos fortificados, con una religión desarrollada que tenía en lugar preminente a un dios del cielo y del trueno que se conserva en el Júpiter que conocemos, con un culto del fuego y con sacrificios de caballos, parece que se dan en un lugar determinado. Gentes que cultivaron el suelo siguiendo de lejos técnicas que se habían desarrollado en el Asia anterior y que para la ganadería, y especialmente para el caballo, recibieron impulsos de las amplias llanuras de Eurasia, alcanzaron esa combinación de agricultura y ganadería que E. Wahle ha señalado como peculiaridad de los indoeuropeos. Los túmulos (*kurgany*) que se encuentran ya en la segunda mitad del IV milenio a. de C. en Ukrania y más al este, hasta la desembocadura del Volga, serían de estos indoeuropeos primeros, en cuyas tumbas tenemos vehículos tirados por caballos.

Una estudiosa de origen lituano, Marija Gimbutas, que ha presentado en los Estados Unidos síntesis en que utiliza los trabajos arqueológicos realizados en la Unión Soviética, Polonia, Yugoslavia, Rumania, etc., ha defendido resueltamente la tesis de que esa cultura de los *kurgany* corresponde a la buscada patria de los indoeuropeos. Desde allí, hacia mediados del III milenio, se pone en marcha una serie de movimientos que hacia el sur, el oriente y el occidente impusieron el predominio de aquellas gentes y de aquellas lenguas, que se fueron diferenciando.

Creemos que las exageraciones racistas no deben hacer olvidar que esos "indoeuropeos" hablaban una lengua. La tesis de que en medio de un caos formativo y mezcla de diferentes tipos lingüísticos resultara al fin una especie de combinación de corrientes varias que podían proceder por ejemplo de las lenguas finougrias que se pueden suponer al nordeste y de las caucásicas al sudeste, para formar el indoeuropeo, fue formulada por admirables lingüistas: C. C. Uhlenbeck (1935) y el Príncipe Trubetzkoy (1939), pero parece poco realista negar la existencia de una verdadera lengua que se refleja, al menos como impuesta a otros pueblos por aristocracias dominantes, en las lenguas que hemos enumerado y que se extendieron desde la India de los antiguos libros védicos de la sabiduría hasta la Irlanda de las leyendas épicas y el derecho medieval.

Ciertamente no nos es conocido el indoeuropeo del 3000 a. de C., pues sólo tenemos textos de lenguas indoeuropeas escritos en el milenio II (hetita y griego de las inscripciones micénicas), o incluso, aunque sobre la base de largas tradiciones orales, en el I (sánscrito de los Vedas, Homero), pero las especulaciones de ciertos indoeuropeístas actuales como W. Meid (1975) distinguen un "indoeuropeo tardío", de la época inmediatamente anterior a la separación de las lenguas que emigraron las primeras (hetita, indoiranio, griego), alrededor quizá de 2000 a. de C. (o antes), frente a otras etapas de "indoeuropeo medio" y "temprano", que se perfilan como hipótesis cada vez más alejadas de nuestras posibilidades de conocimiento.

Un indoeuropeo común puede, pues, hasta cierto punto, reconstruirse sobre la base de las lenguas que han llegado a nosotros ya separadas y en épocas muy tardías, a veces tan tardías como el lituano y las demás lenguas del grupo báltico, bastante conservador y con rasgos muy indoeuropeos en la gramática y el léxico, aunque los textos más antiguos son del siglo XVI, redactados con fines religiosos por predicadores luteranos.

La reconstrucción del indoeuropeo permite reconocer una fonética con correspondencias regulares que dan solidez a la etimología; el acento parece que era musical, lo que dio estabilidad durante largo tiempo en muchas ramas a las vocales. También descubre la reconstrucción una conjugación que en muchos puntos se ha conservado en el sánscrito más antiguo (el védico) y en el griego clásico, si bien ciertos rasgos, que en esas lenguas faltan del todo, o casi, como la voz media o pasiva con *r*, que tenemos en latín y otras lenguas marginales, han de ser tenidos en cuenta; también podemos reconstruir una declinación muy compleja y típica, con los conocidos paradigmas para los diferentes temas, y con un número corto de casos, aún más reducidos en el plural y aún más en el número dual. En la sintaxis el indoeuropeo, a juzgar por las lenguas más antiguas y conservadoras, colocaba el determinante (genitivo, adjetivo, demostrativo) delante del determinado, y por eso iba el verbo en la oración aseverativa al final, detrás de su objeto. En cuanto al vocabulario ya hemos visto que refleja la unidad primitiva en ciertos aspectos culturales de gran significación. Son además comunes los numerales, designaciones de parentesco, nombres de partes del cuerpo, ciertas instituciones religiosas y sociales muy primitivas, ciertas fórmulas poéticas; naturalmente es en las lenguas que se conocen en época más antigua, o en las culturas más tradicionales, donde se conservan más visibles restos de este patrimonio común.

La expansión de los indoeuropeos

Pero la atención fundamental que los comparatistas conceden a estos elementos comunes no debe hacer olvidar que la enorme expansión geográfica del indoeuropeo a lo largo de cinco milenios ha supuesto mezclas e influencias de otras lenguas, que en la mayoría de los casos no nos son conocidas. La idea de lenguas mixtas suscita siempre actitudes apasionadas, y hay posiciones extremas, desde la de los que por razones de método científico niegan que pueda existir una lengua mixta, hasta los que piensan que toda lengua es mixta, pues no ha podido vivir aislada.

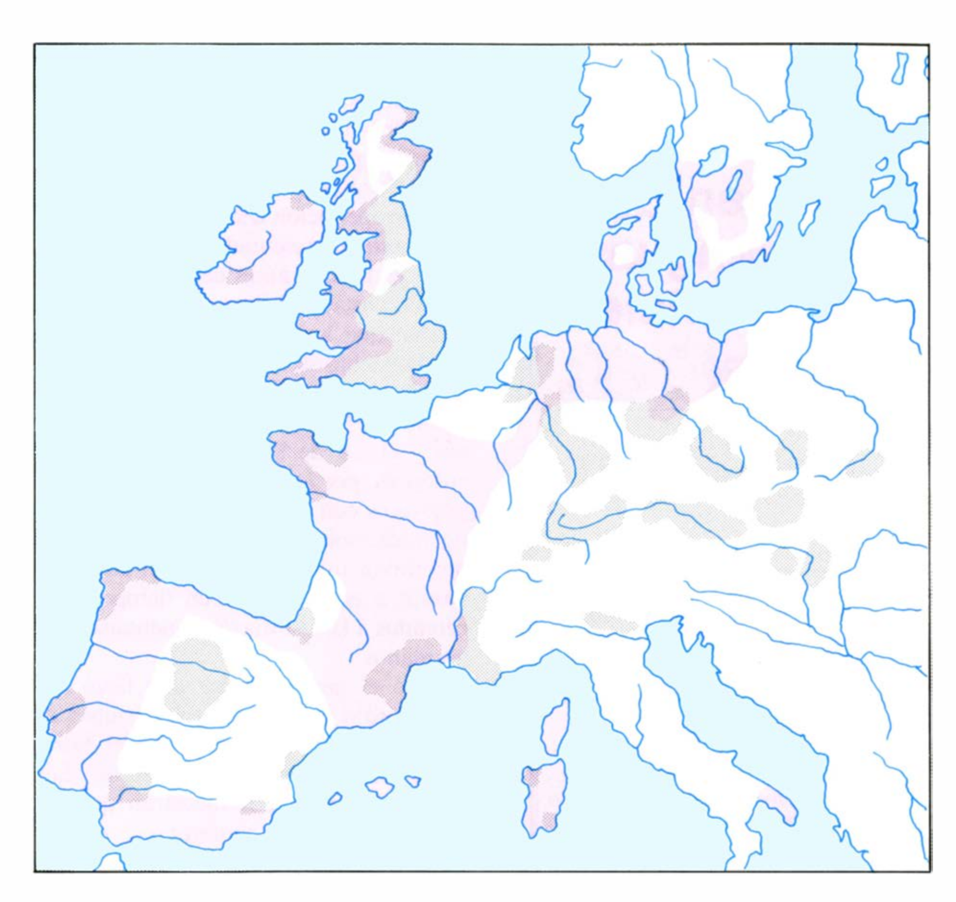
Hay que tener en cuenta que el proceso mismo de la expansión indoeuropea debió ser muy complejo, y así es probable que la difusión de la agricultura, igualmente desde Asia, a través de los Balkanes y la cuenca del Danubio, reflejada en culturas diversas, como las denominadas de Vinča, de Sesklo, de Vučedol, de Cucuteni-Triploje, y otras que subsistieron hasta el milenio III, matizara de un modo muy complicado la difusión de los indoeuropeos hacia el centro de Europa. P. Bosch Gimpera, en su importante libro *El problema indoeuropeo,* combinaba ideas con las que él se había familiarizado junto a Kossina, el defensor en Berlín hasta hace medio siglo de la tesis nórdica, con el estudio de los resultados de excavaciones posteriores, y atribuye a culturas del Danubio y de Europa central en la formación de los indoeuropeos importancia comparable a las de las gentes de Ukrania a que nos hemos referido, y que él llama ponticocaucásicos.

Desde luego la difusión de los indoeuropeos supuso algo nuevo: la sustitución de un complicado mosaico de lenguas por una gran familia. Esto dependió de las posibilidades económicas que hemos señalado: la combinación de una agricultura de cierto desarrollo con una ganadería muy avanzada.

Creemos que puede ayudar a comprender lo que ocurrió en el Antiguo Continente entre el III y el I milenio a. de C. lo que encontramos en América del Sur. Allí tenemos tres familias lingüísticas ampliamente difundidas: los arahuacos, desde las Antillas hasta el trópico de Capricornio, los tupi-guaraní, desde Paraguay hasta la costa atlántica y las

laderas orientales de los Andes en Bolivia y Perú, y los caribes en las costas del norte, desde la vertiente colombiana del Pacífico hasta casi las bocas del Amazonas. Fue su superioridad como agricultores la que les permitió imponerse en el complicadísimo mosaico de las otras lenguas que a centenares se extienden por el continente. Naturalmente que las grandes lenguas no llegaron a formar zonas continuas sino que, según sus hablantes se extendieron, principalmente en canoas a lo largo de los ríos y de las costas, iban dejando a las otras lenguas en los territorios menos favorecidos.

La difusión de estas tres grandes familias permite, en un ambiente muy distinto, comprender la difusión de los indoeuropeos en el Antiguo Continente. La superioridad económica significa superioridad guerrera y cultural, y en su rápida difusión estas culturas de nómadas o seminómadas difunden ciertos progresos de las culturas más altas: las del Próximo Oriente en el Antiguo Mundo, las de los Andes y Mesoamérica en el Nuevo. Tenemos como tres escalones bastante paralelos: las culturas más avanzadas, ya con instituciones estatales y una religión sumamente desarrollada, los grandes pueblos bárbaros, conquistadores y guerreros, y por debajo los pueblos de cultura inferior, aún sin agricultura o casi, necesitados por lo mismo de espacios amplísimos para subsistir, y perdidos en ellos, con pequeñas lenguas muy diversas y aisladas. Tanto los pueblos de cultura superior como los conquistadores en el estadio intermedio son enemigos de los pueblos más atrasados, y los combaten y dominan. Lo mismo los tupí del Brasil llamaban *tapuyas* a la gente de las tribus aisladas, que los peruanos del imperio incaico los llamaban *aucas*: son los despreciados "enemigos", los "salvajes".



EL OCCIDENTE PREINDOEUROPEO que a lo largo de, podemos suponer, casi un milenio (aproximadamente desde antes de 2000 a antes de 1000 a. de C.) resistió a las invasiones indoeuropeas, representa un mundo lingüístico casi del todo desconocido. La tipología lingüística del vasco es muy diferente de la que parece aflorar en el céltico insular (irlandés o galés). Ello indica la variedad de ese mundo sumergido. En el mapa (que trazamos a base de S. Piggot) se señala la extensión de los megalitos (*zonas coloreadas de rosa*) y la del vaso campaniforme ligada a la difusión de la metalurgia, probablemente a partir de Hispania (*zonas punteadas de gris*). El vasco se hablaba por casi todo el norte.

En el mundo que nos interesa, la Europa occidental, el problema de la difusión de los indoeuropeos a costa de otros pueblos y lenguas tiene muchas incógnitas, pues todas las lenguas preindoeuropeas, excepto el vasco, que sabemos se hablaba todo a lo largo de los Pirineos hasta el Mediterráneo, y tal vez también se habló en Cantabria y Asturias, han desaparecido. Es imposible saber nada de todas las otras lenguas de la Europa occidental primitiva. Sólo en el sur, en parte porque las invasiones ya sabemos que llegaron del nordeste, y en parte porque la cultura, con la escritura, llegó antes allí desde el Próximo Oriente, se levanta un poco el velo.

Aparte del vasco, único superviviente, tenemos testimonios del ibérico, lengua de todo el Este y parte del Sur de España como también de la costa de Francia hasta Montpellier. El ibérico, y como él el tartesio de la zona meridional de la Península, no eran tampoco indoeuropeos. En Italia los etruscos no eran indoeuropeos, y es posible que en las montañas de Liguria o de la antigua Retia (Tirol y zonas limítrofes) tengamos en la toponimia o en algunas dificilísimas inscripciones la huella de lenguas preindoeuropeas. En la época clásica había todavía en el interior de Creta o de Sicilia restos de poblaciones primitivas que conservaban lenguas preindoeuropeas.

Pero todas estas lenguas extintas y no descifradas, y cuyos restos en inscripciones en general breves, obscuras y monótonas (o tal le parecen a nuestra ignorancia), o en topónimos o palabras supervivientes en diferentes lenguas vivas, no dan sino una idea muy limitada, no hacen sino llamar la atención hacia una realidad misteriosa y poco accesible.

Sin embargo, que las lenguas indoeuropeas, en su difusión hacia el oeste, quedaran matizadas por ellas, al quedar conquistados o simplemente asimilados quienes las hablaban, es una hipótesis muy sólida. Las lenguas célticas sobrevivientes, que llegaron a Gran Bretaña e Irlanda en una época muy tardía, son muy divergentes, en el orden de palabras y en otros rasgos de su gramática, de cualquier otra lengua indoeuropea. Y a la tesis de la "pureza racial", que las doctrinas racistas atribuían a los germanos, oponían algunos lingüistas, no sin malignidad, la explicación de que el cambio completo, llamado rotación, de las consonantes en germánico (que se da bastante semejante en el armenio, rodeado de exóticas lenguas caucásicas), por el que a la *p* de *pater* corresponde con regularidad la *f* de *father*, a la *d* de *decem* la *t* de *ten*, y la *dh* que tenemos en otras lenguas, la simple *d* del inglés *red* 'rojo', pudiera muy bien haber resultado de la indoeuropeización de los nórdicos por invasores llegados de otra parte.

No se puede desconocer que la indoeuropeización, portada por grupos de conquistadores, o a veces de gentes que no hallaron resistencia y se instalaron en medio de pueblos abiertos a la mutua influencia, aunque al principio se impuso como organización social, como religión y como desarrollo económico y cultural, hubo de recibir también el sello

de los sometidos o asimilados, y posiblemente en medida cada vez mayor, sobre todo en las zonas más lejanas, como las Islas Británicas o la India.

Parece podemos deducir de los resultados de la arqueología dos etapas en la indoeuropeización del Occidente. La primera está representada por la penetración, ya mucho antes del año 2000, hacia las costas del Mar del Norte de un pueblo que los prehistoriadores llaman de las hachas de combate. Parece que la constitución étnica de esos territorios de Jutlandia, Holanda, el norte de Alemania y el sur de Escandinavia, adquiere entonces, por la mezcla de los invasores guerreros con los agricultores herederos de tradiciones occidentales, los perfiles definitivos que la caracterizan aún. Ahí vamos a encontrar ya en tiempos más cercanos a la historia documentada a los germanos.

Pero el gran impulso que llevó tan temprano a las gentes de las hachas de combate a las orillas del Mar del Norte tardó largo tiempo en repetirse. Los mapas arqueológicos muestran que los indoeuropeos no continuaron en seguida progresando, sino que todavía hacia 1500 a. de C. sus avanzadas se encontraban en el norte de la actual Francia, entre el Rhin y el Sena. Las nuevas oleadas que llegaron hasta el Atlántico y ocuparon Irlanda y Portugal esperaron aún varios siglos.

El desbordamiento sobre el Occidente extremo

El Occidente preindoeuropeo que a lo largo de, podemos suponer, casi un milenio (aproximadamente desde antes de 2000 a antes de 1000) se mantuvo frente a las invasiones indoeuropeas, representa, como hemos dicho, un mundo lingüístico desconocido y que sería muy audaz caracterizar. Era sin duda muy variado, como parece enseñarnos la tipología lingüística, y debemos imaginar que no fue meramente pasivo frente a la expansión indoeuropea. En los nombres de ríos, a que nos vamos a referir en seguida, también entraron a formar parte elementos lingüísticos procedentes de ese primitivo Occidente, en el que se señalaron sucesivamente las culturas megalíticas (quizá durante dos milenios a partir de 3000 a. de C.) y las que se caracterizan por el vaso campaniforme [Bell Beakers, Glockenbecher-Kultur] (dos o tres siglos antes de 2000 y otros tantos después).

Para caracterizar de algún modo ese antiguo Occidente nos limitaremos a un par de rasgos lingüísticos, meros indicios de una realidad muy poco accesible: la palabra 'nombre' en indoeuropeo y urálico, y el nombre de la plata que tenemos en vasco y en las lenguas germánicas y eslavas.

Es una cosa bien sabida (por ejemplo el conocido especialista sueco B. Collinder insistió en ello) que la palabra indoeuropea para 'nombre' (latín *nomen,* inglés *name,* griego *ónoma,* sánscrito *nama,* irlandés *ainm,* etc.) tiene exacta correspondencia en las lenguas finougrias (finlandés *nimi,* húngaro *név, neve,* yurako *nim,* vogul *näm,* etc.), lo que parece significar que la frontera nordoriental de las lenguas indoeuropeas no estaba cerrada. A ello podemos contraponer que la Europa occidental parece tenía sus puertas abiertas al mundo de las lenguas camito-semíticas: la palabra vasca para 'nombre', *izen,* encuentra parecidos evidentes en Africa y Asia (bereber *isem,* árabe *ism,* nubio *esmi, esim, esum,* etíope *sem,* etc.; en egipcio antiguo para 'nombrar' se halla el verbo *smj*).

Si en esta palabra parece tratarse de una comunidad que supone origen idéntico o relaciones muy antiguas, la otra en que nos vamos a detener un momento es una típica palabra cultural viajera. Para designar la 'plata' en asirio se halla *ṣarpu* o *ṣarpi,* participio de un verbo que significa 'acendrada, purificada al fuego'. Junto a una forma más divulgada, *fidda,* hallamos en árabe la misma palabra que en asirio: *ṣarif* 'plata'. Igualmente tenemos en bereber *azref, azerfa,* y ya al sur del Sahara, en el África negra, hausa *azurfa.*

En vascuence hallamos para 'plata' *zil(h)ar, zillar, zirar, zidar.* Es muy probable que en el bronce celtibérico de Botorrita, hallado hace poco cerca de Zaragoza, signifique lo mismo la palabra *silabur,* y parece razonable que *śalir,* que se lee en ibérico en monedas de plata, denote el mismo metal. Las tres formas de la Península, la vasca de significación segura, las otras dos, probable, se apoyan mutuamente para formar el necesario puente entre las formas con *a* en la primera sílaba (que llegan hasta el ibérico) y las que aparecen con *i* como en vasco (y celtibérico) en las lenguas germánicas y baltoeslavas, para las que hasta ahora no se ha dado ninguna explicación etimológica: gótico *silubr,* inglés antiguo *siolfor, seolfor,* alto alemán *sil(a)bar,* lituano *sidabras,* eslavo antiguo *s(ĭ)rebro,* etc. La divergencia que muestran las formas vasca e ibera respecto de todas las demás se explica en la fonética de estas lenguas, que carecen de *p* y tienen otras peculiaridades en las labiales.

La cronología del inicio del uso de la plata (que se escalona entre los comienzos del III milenario en Asia menor y el

final de la época de bronce en Europa central) parece se ajusta a la difusión de la palabra, pues justamente el puente entre Oriente y Europa central lo formaría la aparición de plata en Almería en los comienzos de los metales, y en un ambiente de influencias orientales. Es bien conocida la importancia de las minas, también de plata, en la Hispania antigua, y la cultura del vaso campaniforme, que muchos autores consideran tuvo en la Península su centro, fue portada por prospectores, mineros y metalúrgicos que dieron lugar a culturas secundarias en muchas zonas, hasta Europa central, Dinamarca y las Islas Británicas. La minería, iniciada en las altas culturas orientales, siguió varios caminos tal vez, pero en el caso de la plata podemos rastrear el que pasó por la Península Ibérica.

La diferenciación de las lenguas indoeuropeas

Si intentamos, pues, sustituir esquemas abstractos por la realidad, procurando entender el marco geográfico y cronológico, nos encontramos con que la idea generalmente admitida de una diferenciación del indoeuropeo simultáneamente en los grandes dialectos no es tan evidente como parece.

Una cosa es la migración, en la época del que hemos visto se puede llamar indoeuropeo tardío, de los hetitas, los indoiranios y los griegos, y otra la división de las lenguas indoeuropeas occidentales, cuyos hablantes eran durante largos siglos nómadas o casi, y no se alejaron tanto de la zona de formación ni entraron en contacto con altas culturas, ni perdieron del todo la vecindad con pueblos afines. W. Meid habló de "bloques residuales postindoeuropeos", es decir, grupos aún poco diferenciados que sin duda se conservaban en el II milenio.

Ya Hans Krahe había señalado hace tiempo que en la época en que el griego, el indio y el hetita eran lenguas perfectamente conocidas, las que luego van a aparecer como lenguas occidentales no estaban del todo separadas ni acaso formadas, y a lo sumo apuntaban como gérmenes en una masa relativamente uniforme que él propuso designar con el nombre de "europeo antiguo" *(Alteuropäisch)*.

Como la escritura no se usaba más que en el Próximo Oriente, no poseemos naturalmente ningún texto de aquellos tiempos del interior de Euopa, pero él se fundó en un sector de la toponimia, los nombres de río, a veces particularmente conservador, para intentar penetrar en aquel mundo lingüístico incógnito. Krahe hizo más rigurosa la recolección y



DERIVADOS DE LA RAIZ INDOEUROPEA **aw(e)-* "mojar, fluir". Como puede apreciarse, se sitúa en puntos más bien centrales de la antigua Europa. Cada número se refiere a un río, del que se da la forma antigua, o la moderna, según la conocemos. Los asteriscos indican formas reconstruidas.

análisis de la hidronimia sobre la que ya habían construido teorías los "iliristas", y antes los "liguristas" (que atribuían a estos poco conocidos pueblos prehistóricos nombres en toda Europa, de Noruega a Sicilia y de los Países Bálticos a Irlanda) y encontró en ella el reflejo de esa época que se puede bien llamar "post-indoeuropea" en la que hay que suponer terminan por cristalizar las lenguas occidentales.

La naturaleza especial de estos materiales onomásticos no da sin duda más que una visión limitada, y seguramente forzada en cuanto a su unidad. Por ejemplo, la frecuencia en estos nombres de la vocal *a*, que en indoeuropeo ocupaba una posición especial, fuera de la alternancia, plantea un problema difícil. Pero el hecho es que *Ala, Ara, Arga, Ava, Pala, Sala, Tara*... son nombres que se repiten, bien en estas formas primarias, bien en derivaciones que se pueden agrupar en series con cada raíz: *Alara, Alanta, Alentia..., Arva, Aranta, Arantia*... Ofrecemos en dos mapas ejemplos de la difusión de tres de estas raíces distribuidas de manera diversa.

Krahe sostuvo que estos hidrónimos de la antigua Europa eran todos indoeuropeos puros; nosotros creemos, por el contrario, que se trata de series incorporadas a un sistema indoeuropeo de derivación, pero las raíces son unas veces de etimología indoeuropea, y otras en cambio han sido incorporadas desde otros mundos lingüísticos, especialmente desde el desconocido de Occidente.

El primero de los mapas que damos en estas páginas presenta derivados de la raíz indoeuropea **aw(e)-* 'mojar, fluir'; en el otro hemos recogido derivados de

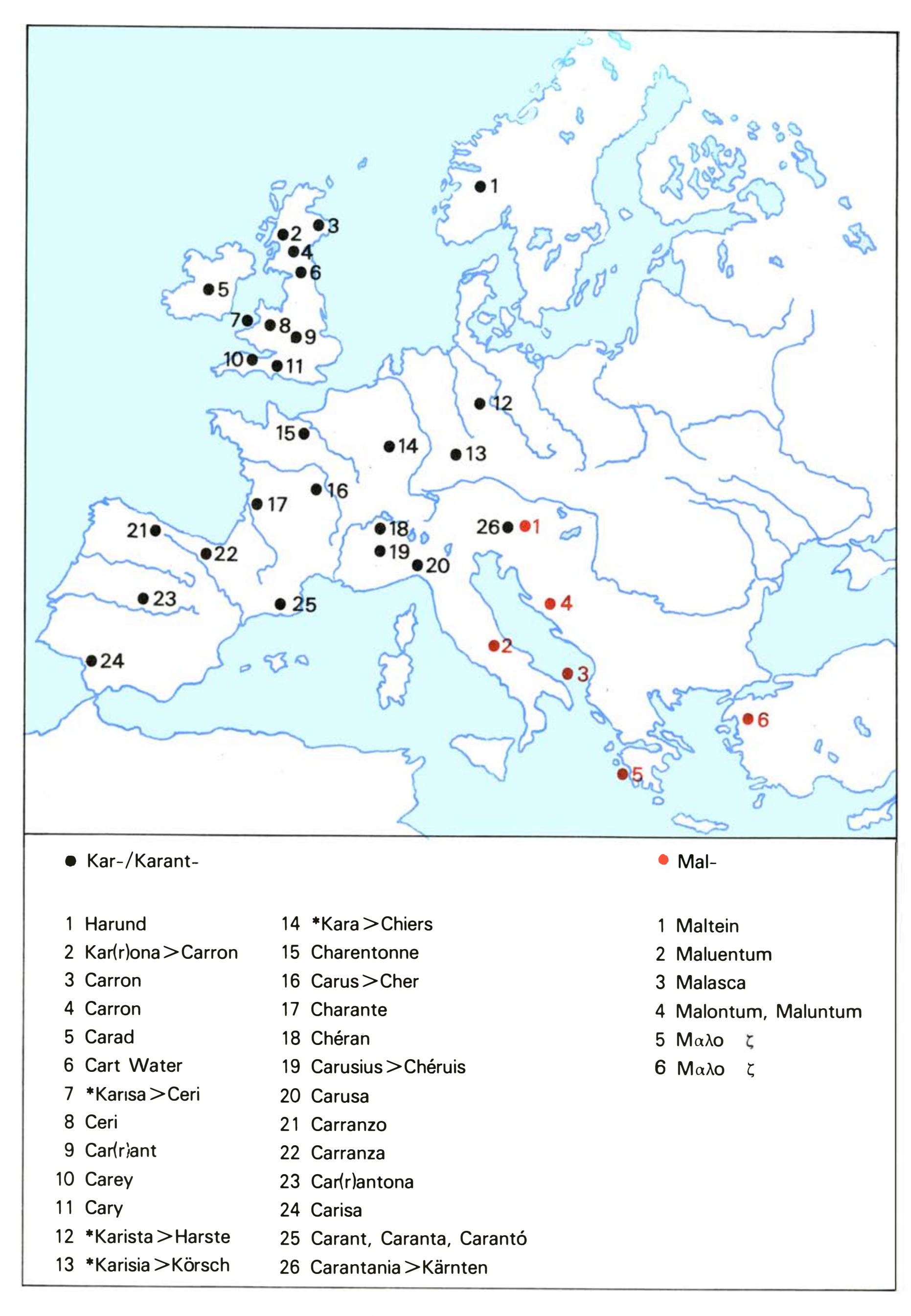
**kar(r)a* y **mal-*, que significan 'piedra' y 'roca' y no son de etimología indoeuropea: baste para el primero recordar el vasco *(h)arri* 'piedra' o el irlandés *carrac* 'roca, arrecife' o muchos nombres de lugar de los que recordaremos *Carrara,* famosa por su mármol. Una simple ojeada permite comprobar en los mapas que la distribución de la primera raíz ocupa zonas más centrales, mientras que las otras son claramente marginales.

Los hidrónimos de la vieja Europa son, pues, testigos de una época en que las fronteras lingüísticas eran muy abiertas y los habitantes de la Europa bárbara, alejados de las altas culturas, eran más bien nómadas.

Se puede bien partir de que una división simultánea y simétrica del indoeuropeo en los grandes dialectos o lenguas no ocurrió. Detrás de los primeros pueblos que abandonaron la "patria común", quedaron grandes grupos de nómadas en expansión, en vastos territorios donde pueblos y tribus se movían, se mezclaban y separaban, absorbían a pueblos sometidos o vecinos, se convertían en centros de poder o de cultura o religiosos, o, por el contrario, eran absorbidos y desaparecían.

Tal fue durante quince o veinte siglos la situación en Europa central y, comenzando más tarde, en la occidental.

El celta o el germánico, grupos lingüísticos a que nos limitamos aquí, pues el problema del latín en Italia es distinto y exigiría otro estudio extenso, no se formaron contemporáneamente. Surgen no sólo en lugares distintos, sino en tiempos separados, con historias primitivas muy diferentes y como cristalización de procesos que sólo alcanzamos a comprender esquematizándolos.



DERIVADOS DE LAS RAICES PREINDOEUROPEAS **kar(r)a* y **mal-*, que significan "piedra" y "roca". Como en el mapa de la página precedente, los números indican la situación de los ríos que van numerados en la lista. La distribución de estas formas es más marginal que la del mapa anterior.

El celta

Según una idea algo simplificada, el celta se manifiesta con sus rasgos característicos en dos grupos de lenguas: uno, formado por las marginales de Irlanda (goidélico) y de Hispania (celtibérico), ha conservado el sonido *qu* del indoeuropeo que vemos por ejemplo en la conjunción enclítica *que,* en latín y en celtibérico con la significación de 'y'; el otro grupo de lenguas celtas se manifiesta con una diferenciación que tuvo su centro de formación con la llamada cultura de La Tène en las regiones del este de Francia y sur de Alemania, y presenta la evolución de *qu* a *p;* por ejemplo la palabra *petorritum,* un carro de cuatro ruedas de los galos, es un compuesto del numeral 'cuatro' (latín *quattuor*), que hallamos también en galés *pedwar* 'cuatro'.

Los pueblos celtas que toman la facies cultural de La Tène aparecen inmediatamente como galos históricos, que se extendieron también por Gran Bretaña, donde predominó un dialecto semejante al suyo, y por Italia, y llegaron más tarde, ya en el siglo III a. de C., a Asia Menor. Pero tenemos que corregir la idea, que predomina, de identificar a todos los celtas con los galos. Los celtas son anteriores a la formación de la cultura de La Tène, y ésta no significó en ninguna parte una ruptura con la tradición anterior. Lo que ocurre es que, al irse remontando en el tiempo, la identificación de pueblos y culturas arqueológicas se vuelve más problemática. Podemos estar seguros de que en la primera Edad del Hierro, la que se llama cultura de Hallstatt, tuvieron papel esencial los celtas, pero ya es más difícil equiparar época de Hallstatt con formación de los celtas.

Las invasiones que llevaron a los celtas a Irlanda y a Hispania ocurren en épocas difíciles de determinar, al menos con los datos que combinan los arqueólogos. Sin embargo, en un gran trabajo monográfico de W. Schüle hallamos una identificación resuelta, que a nosotros nos parece muy plausible, entre los celtíberos y la que él llama cultura del Tajo, cuyos rasgos encuentra fijados en el siglo VI. Anterior y con otros componentes es la cultura que el mismo prehistoriador llama del Duero, y que nosotros propondríamos, desde el punto de

vista de la lingüística, que se pusiera en relación con el lusitano.

La celtización de las Islas Británicas es un problema más difícil, por quedar, sobre todo Irlanda, demasiado alejadas del mundo conocido por las culturas antiguas. El prehistoriador Chr. Hawkes ha calificado de "celticidad cumulativa" las repetidas migraciones de grupos pequeños o grandes que, podemos suponer, unas veces como guerreros, otras de modo menos violento, fueron transformando el panorama arqueológico de las Islas, a partir por lo menos de los finales del bronce y de una manera indudable con la presencia de la cultura del Hallstatt en el siglo VII. Irlanda conservó, como ya hemos señalado, una forma más arcaica, el goidélico, mientras que en la Gran Bretaña, con migraciones históricamente comprobables de gentes de la cultura de La Tène desde Francia, se desarrolla un celta que en la época de César era sin duda inteligible para los galos de Francia.

Se ve, pues, que la celtización del Occidente, hasta los más remotos rincones de Gran Bretaña e Irlanda, hasta la Celtiberia de las cabeceras del Duero, el Tajo, el Júcar y el Turia, hasta Galicia y el país de los Célticos en el Alentejo, es un fenómeno muy complejo, cuyas primeras oleadas se señalan ya en la época de los campos de urnas, en el siglo VIII.

Queremos sacar algunas consecuencias de un hecho lingüístico quizá no bastante tenido en cuenta: las lenguas indoeuropeas distinguen en general tres clases de consonantes oclusivas: sorda, sonora y aspirada, que se pueden representar esquemáticamente *p b bh, t d dh, k g gh* y k^w g^w gh^w. Ahora bien, en la zona central de su difusión las lenguas indoeuropeas redujeron a dos clases estas consonantes, dejando la aspiración. Las lenguas iranias, que en la antigüedad ocuparon durante siglos (principalmente los escitas hablaron un dialecto iranio) los territorios del norte del Mar Negro, quizá fueron el centro de esta reducción. El sánscrito conserva las aspiradas. Pero el eslavo y el báltico, como el albanés, han perdido las aspiradas igual que el iranio. Igualmente las han perdido otras lenguas más o menos hipotéticas de Europa central, como el ilirio y el ligur, y también las formas de la hidronimia antigua carecen de aspiradas. A esta misma clase de lenguas sin aspiradas pertenecen las lenguas célticas, mientras que las lenguas marginales: el germánico al norte, y al sur el latín con otras lenguas itálicas, y el venético, así como el griego y el armenio, presentan en formas distintas las tres clases de consonantes oclusivas.

Ese rasgo del céltico, como del "antiguo europeo" y de otras lenguas que se desarrollaron en el centro (incluyendo en ellas al frigio, pariente del tracio y llegado ya en el I milenio a Asia Menor), podría pensarse que proviene de las incursiones de pueblos jinetes que desde el fondo de Asia y desde las llanuras de Europa oriental aparecen como componente, según acreditan ciertos detalles de los arreos de caballo que los arqueólogos encuentran desde China hasta Portugal, de los movimientos de pueblos indoeuropeos. En la cristalización de los celtas un componente de este tipo pudo ser decisivo.

El germánico

La dificultad que la lejanía de los centros de la cultura antigua representa, en mayor grado que para la Península Ibérica para las Islas Británicas, ocurre también para el problema de la formación de los germanos. Ya hemos visto que los componentes que entraron en la formación de los germanos se puede considerar que datan del III milenio con la llegada de los pueblos de las hachas de combate, llamados también de enterramientos individuales *(Einzelgrabkulturen),* al ambiente agrícola occidental primitivo. Pero los arqueólogos no se deciden a identificar plenamente a los germanos con sus rasgos propios sino mucho más tarde: poco antes de los mediados del I milenio a. C. (R. Hachmann). Aparecen entonces culturas materiales en las costas del Mar del Norte y del Báltico las cuales se extienden hasta el sur de Noruega y Suecia y hasta los confines meridionales de la llanura de Alemania, en las cuales tenemos ya a los germanos históricos.

Lo que podemos saber de la difusión del germánico es sumamente importante para la difusión de los dialectos indoeuropeos de formación más antigua. En el caso del germánico la historia nos permite distinguir una cristalización antigua y una difusión posterior, la cual podemos ver que fue complicada por causas históricas y culturales que en cierta medida podemos conocer.

Ya hemos visto que en la Europa central, y más tarde en la occidental, dialectos indoeuropeos fueron portados por diversos pueblos y tribus que en su expansión se influían y mezclaban, se absorbían mutuamente y se configuraban a consecuencia de la formación de centros de predominio político, económico, religioso, etc. Lo que no tenemos derecho es a extrapolar un milenio o dos hacia atrás pueblos y culturas que nos son conocidos mucho más tarde. Los pueblos y lenguas que aparecen más tarde se habían formado en aquel

	abedul	salmón	lobo	caballo	buey, vacuno	oveja	lana	cerdo
HITITA								
SANSCRITO	bhūrjáḥ	(lākṣā 'laca', dudoso)	vŕkaḥ	áśvaḥ	gáuḥ	áviḥ	ūrṇā	sū-kara-
AVESTICO	(oseta) bärz		vəhrka-		gāuš		varəanā	hū
GRIEGO			lýkos	(h)ippos	boûs	o(w)is	lênos	hŷs
LATIN	farnus 'fresno'		lupus	equus	bōs	ouis	lāna	sūs
OSCO Y UMBRO					(u.) bum (acusativo)	(u.) oui		(u.) sim (acusativo)
IRLANDES				ech	bó	ó i		socc sáil 'cerdo de mar, un pez'
GALES				ebol 'potro'	buch	ewig	gwlan	hwch (galo)+su-
GOTICO			wulfs			awi-str 'aprisco'	wulla	
ALTO ALEMAN ANTIGUO	birihha	lahs	wolf	(ingl. ant.) eoh	chuo	ouwi	wolla	sū
ISLANDES ANTIGUO	bjǫrk	lax	ulfr	iōr				sȳr
ESLAVO ECLESIASTICO	(ruso) berëza		vlĭkŭ		* govędo	ovĭca	vlŭna	svinŭ
LITUANO	béržas	lašišà					vìlna	(letón) suvēns 'cochinillo'
TOCARIO		(B) laks 'pez'		(B) yakwe	(A) ko (B) kau			(B) swañana misa 'carne de cerdo'

CUADRO CON EJEMPLOS DE PALABRAS que aparecen en la mayoría o en muy distintas ramas del indoeuropeo. Con ellas se pudo construir la teoría de una "patria común". Los prehistoriadores alemanes y escandinavos se inclinaron a menudo a poner en el norte de Europa la patria de los indoeuropeos. Pero ante la exageración racista de identificar pueblo con lengua, muchos comparatistas se apresuraron a negar que los "indoeuro-

puerco	agro	grano	yugo	cobre, bronce	(Dios del) cielo, día	rey	pueblo	hombre libre
			i-ú-ga-an			rāj–	tuzzi- 'señor, campamento'	
	ájraḥ		yugám	áyas- 'metal, hierro'	dyauḥ 'cielo', devah 'dios'			
(saka) pāsa				ayaṅh- 'metal, hierro'	daeva- 'demonio'			
	agrós		zygón		Zeús, Diós; dí(w)os 'divino'			eleútheros 'libre (no esclavo)'
porcus	ager	grānum	iugum	aes	Iūpiter, Iouis; diēs 'día'; deus 'Dios'	rēx		līber 'libre'; līberī 'los hijos (no esclavos)'
(u.) porca	(u.) ager				(o.) Diúveí (dativo)		(o.) touto, (u.) toutam 'ciudad'	(o.) Lúvfreís (genitivo) 'el dios Liber, Baco'
orc 'cochinillo'		grán			díe 'día', dia 'Dios'	rí, (celtiber.) -reigis	túath	
		grawn	iau		dydd 'día' duw 'Dios'		túd 'país'	
	akrs	kaúrn	juk	aiz		reiks (del celta)	þiuda	
far(a)h	ackar	korn	juch, joch	ēr	(ingl. ant.) Tīg 'Marte' Tuesday 'martes'		diot(a)	liute
		korn		eir	tīvar 'dioses'			
prasę		zrŭno	igo		divŭ 'milagro'			ljudĭje 'gente', ljudinŭ 'hombre libre'
paršẽlis 'cochinillo'							(letón) tàuta	
			[(A) yokäm 'puerta', posiblemente préstamo indio]					

peos" de la etapa primitiva común fueran un verdadero pueblo con una verdadera lengua. En este cuadro se recogen palabras referentes a la ganadería y agricultura, al primer metal útil y a la religión e instituciones. La "revolución democrática" que Meillet supuso entre los indoeuropeos centrales parece reflejarse en las palabras "pueblo" y "hombre libre", mientras que la palabra "rey" sobrevivió en sánscrito, latín y celta (*derecha*).

mundo inestable que nos es difícilmente conocible, y cada lengua y cada pueblo tomaron en un cierto momento, que a menudo nos es imposible determinar, sus rasgos diferenciales. ¿Se puede hablar de celta antes de la pérdida de la *p* indoeuropea? ¿De germano, antes de la rotación consonántica? Creemos que no.

La historia de Europa en la Edad Media nos muestra que la formación de pueblos y lenguas puede ocurrir, determinada por vicisitudes de la historia, de modo bastante repentino. Pensemos que en los vastos territorios donde en el siglo IV se hablaba latín o se hablaba germano, en el siglo X hallamos que se hablaba francés, o castellano, o provenzal, o alto alemán, o sajón antiguo... Y la cristalización de tales lenguas resulta de la absorción de ciertos dialectos y variedades por otros más prestigiosos y centrales.

En el caso del germánico, que según estudios de filólogos clásicos entra en contacto con los romanos en la expedición de los Cimbros y Teutones poco antes del año 100 a. C., se puede ver que la "germanización" avanza hacia el Bajo Rhin en los Países Bajos y hacia el valle del mismo río que forman Alsacia y Baden, no sólo mediante movimientos de tribus, sino también atrayendo a pueblos que parece eran distintos. César, que vio a los germanos como un peligro y justificó así su conquista de Galia, habla de tribus de la actual Bélgica que él llama *Germani Cisrhenani*, es decir, germanos de la orilla izquierda del Rhin. Pero resulta que estos "germanos" llevan nombres que no tienen ningún rasgo germánico. Nuestra interpretación de este y semejantes hechos es que tribus de la gran masa de pueblos indoeuropeos o indoeuropeizados (si es que al cabo de varios siglos tiene sentido tal distinción) eran atraídas por pueblos en proceso de crecimiento, y tomaban con la cultura la lengua de ellos. El historiador suizo G. Walser, que ha estudiado el testimonio de César, encuentra que, al lado de los celtas, los germanos aparecen en los informes del político romano "todavía en un estado como preétnico", sin fijar del todo en cuanto a sus caracteres nacionales, su cultura y costumbres.

A mi juicio en este caso de las fronteras extremas del mundo germánico en expansión parece que las vemos extenderse, y no con invasiones ni migraciones, sino con la adopción del "germanismo" por otras gentes. En efecto,- podemos seguir el avance germánico en la pronunciación del río que ahora se llama Wal, un afluente del Escalda casi en la desembocadura. En César aparece en la forma *Vacalis*, pero siglo y medio más tarde, en Tácito, se llama ya, con la rotación consonántica, *Vahalis*. Ya tenían cuidado los romanos de que los germanos no entraran como dominadores en tierras del Imperio, pero sin duda que pueblos más o menos afines miraban con simpatía la expansión germánica. De otros fronterizos que muestran nombres galos o de esa zona intermedia a que vamos a referirnos, los Tréviros y Nervios, nos dice Tácito que "con orgullo se jactaban de origen germánico"; seguramente se inventaban un origen germánico que sus nombres y su lengua no comprobaban, para dejarse atraer por una lengua en expansión.

Estudiando inscripciones romanas de Rhenania se puede ver el proceso de esta germanización que avanza y se va atrayendo gentes: una dedicatoria a las *Matronae Gesahenae* corresponde con la fonética germana de la *h* al nombre de un dios galo *Gesacus* (y aparecen también las mismas divinidades en la forma no germánica *Matronae Gesationum*). La dedicatoria *Matronis Iulineihiabus* procede de la ciudad de *Iuliacum*, hoy Jülich. La misma declinación germánica surge tímidamente en inscripciones donde predominan formas celto-latinas, por ejemplo tenemos *Matronis Vatvims* en germánico frente a la forma más normal del dativo *Vatviabus*.

La absorción de lenguas menores

Lo que la historia nos permite ver en la difusión y cristalización de celtas y germanos es, pues, no el final de un desarrollo esquemático en el cual los indoeuropeos se dividieron desde el principio en los dialectos conocidos, sino el resultado de un desarrollo azaroso y complicado en el que la cristalización de grandes lenguas y desaparición de otras resulta de las múltiples vicisitudes y azares de la historia.

El germánico aparece, como se deduce de los trabajos de un gran filólogo clásico, E. Norden, poco a poco a la luz de la historia. Los nombres de los jefes de Cimbros y Teutones, y los de estos mismos pueblos, todavía no fueron transcritos como sonarían en germánico. En César, medio siglo más tarde, sí que tenemos atestiguados nombres germánicos: los *Suebi, Cherusci, Marcomani, Vangiones*, formas que ya muestran los rasgos característicos. Pero, como nos dejan ver César y Tácito, la expansión de estas características germanas parece gradual, y diríamos que podemos percibir que fueron aceptadas voluntariamente en muchos casos. Se diría que no necesitaron pasar el Rhin muchos germanos para que al fin se decidiera la germanización del territorio de Flandes.

Muchos estudiosos de la toponimia de estas regiones de las orillas del Mar del Norte han desarrollado la hipótesis de que en ellas hay formas características, que no son célticas ni germánicas, que a veces parecen contener elementos preindoeuropeos, y que forman lo que ellos llaman "bloque del Noroeste" (nos limitaremos a citar los trabajos del germanista H. Kuhn). Según tal explicación, estos pueblos de los Países Bajos y zonas limítrofes hasta el río Weser, en la expansión de los dos grandes grupos de celtas y germanos, fueron al fin atraídos por éstos, si bien tal proceso se debió encontrar frenado por la política de temor al peligro germánico que César previó y sus sucesores en el solio imperial mantuvieron.

Esta difusión de una lengua mayor atrayendo, o imponiéndose, a tribus o grupos lingüísticos menores, tal como sobre la base poco segura de los topónimos la vemos en esa región del Noroeste de Germania, aparece un poco más clara en la Península Ibérica. Tenemos bien identificada una lengua céltica de tipo no galo en Celtiberia, con monumento tan extenso como el bronce de Botorrita y con una serie de textos epigráficos tanto en escritura ibérica como latina. Pero precisamente en la región entre el Duero y el Tajo, alrededor de la Serra da Estrela, tenemos tres inscripciones en otro idioma indoeuropeo que podemos sin duda llamar lusitano. Esa lengua, de la que sólo unas pocas palabras se identifican con toda seguridad, se diferencia del celta en uno o dos rasgos fundamentales: la *p*, que se halla en muchas formas "preceltas" del Occidente de Europa, y el tratamiento de las aspiradas, que si tenemos razón en la interpretación de una palabra de la inscripción del Cabeço das Fráguas, no sería como en celta, sino como en las lenguas marginales.

La conservación del lusitano, lengua que se puede llamar "precelta" o "paracelta", enseña que las que aparecen como grandes lenguas o familias de lenguas de Occidente, así el celta o el germánico, son el resultado de una cristalización que desde el principio llevó consigo complejos procesos de absorción y fusión. En la Hispania antigua hubo sin duda otros dialectos indoeuropeos que posiblemente tendieran hacia el polo del celtibérico, o hacia el polo del lusitano, pero que no son otra cosa que grupos lingüísticos menores, voluntaria o forzosamente siempre en camino de incorporación a una lengua mayor. Las grandes lenguas absorbieron otras.

Ciencia y sociedad

Psicología de la educación: una ciencia a debate

De todas las ramas de la psicología, la psicología de la educación es, en el momento actual, la que probablemente más controversia va a despertar en los próximos años y más trabajo va a exigir de sus profesionales. Y esto es así porque si bien su "ranking" entre las ciencias del comportamiento va sin duda en aumento, la misma idiosincrasia de su constitución –a caballo entre dos campos, la psicología y la educación– la hace aparecer como extremadamente vulnerable a diferentes tipos de problemas. Por ejemplo, y para empezar, el mismo hecho de su definición. En principio es aceptable considerar a la psicología de la educación como aquella disciplina que trata del sujeto que aprende, del proceso del aprendizaje y de las estrategias instruccionales necesarias para que se cumpla la educación. Nada habría de objetable a esta definición si no fuera porque coexiste con una docena más de definiciones parecidas o totalmente distintas entre sí. Así, las hay que sólo incluyen como objeto de la psicología de la educación el proceso del aprendizaje y la retención, otras el estudio científico de la práctica educativa, otras que la consideran como simple psicología general aplicada a la educación, otras como una prolongación de la psicología del desarrollo. Otras como un capítulo de la psicología social, etcétera.

En relación con los marcos teóricos de referencia necesarios para una explicación científica válida tampoco se caracteriza la psicología de la educación por su unidad. Ciertamente que la pauta temática establecida por su supuesto fundador –Edward L. Thorndike (1874-1949)– ha pervivido hasta nuestros días, y el hecho de considerar a la psicología de la educación como una ciencia del mismo estilo que las ciencias físicas ha beneficiado, cuando menos, a algunas de sus mejores investigaciones, por ejemplo, en el campo de la medida del rendimiento. Sin embargo, junto a este modelo, psicólogos sociales, interaccionistas, funcionalistas, cognitivistas, "humanistas", e incluso innovadores con intención de sustituirla por una nueva ciencia, y naturalmente por un no muy afortunado nuevo nombre, Educología, se reparten un campo que es, en muchos aspectos, tierra de todos y de nadie. Y junto a la falta de una definición homogénea y un modelo explicativo coherente no es menos cierto también, y probablemente a causa de ello, que la psicología de la educación tampoco posee aún su propia historia escrita.

Ante este panorama, cuando menos sorprendente, de una ciencia cuya estructura se suponía consolidada hace mucho tiempo, vale la pena preguntarnos si no hay algo que decir, aunque sea idealmente, menos sombrío que lo hasta ahora señalado. Ciertamente, aparte de la definición tentativa más arriba apuntada, se pueden señalar algunos rasgos orientativos sobre la psicología y el psicólogo de la educación. En primer lugar resulta bastante coherente afirmar que la psicología de la educación trata del análisis sistemático de la dinámica del comportamiento del escolar, de las características de la enseñanza efectiva y del medio ambiente que le rodea, lo que se ha dado en llamar últimamente la ecología de la clase. En segundo lugar se pueden abstraer, de entre todas las teorías aproximativas existentes, algunos objetivos válidos y que realmente le son propios: aplicación de los presupuestos psicológicos al proceso educativo, diseño de una experimentación adecuada que contribuya a mejorar el conocimiento de los procesos implícitos en el binomio "enseñanza-aprendizaje" y, finalmente, el desarrollo de teorías y modelos representativos de lo que "realmente" ocurre en la clase. En tercer lugar, aquellos temas que conciernen, al menos en la sociedad de hoy, al profesor, al alumno y a la cultura y que son: (*a*) el desarrollo del niño y lo que ocurre, instruccionalmente hablando, durante el mismo, por ejemplo, los procesos de socialización; (*b*) los procesos de aprendizaje, por ejemplo lo que hace referencia a un programa instruccional efectivo; (*c*) la medida y evaluación del progreso escolar, por ejemplo los problemas del rendimiento y las deficiencias en el aprendizaje, y (*d*) la higiene mental en la clase, por ejemplo las relaciones buenas o menos buenas entre alumnos-alumnos o profesor-alumno. Finalmente, se debe añadir que el psicólogo de la educación es un profesional que puede cuidar, y de hecho así lo hace en algunos países, de la formación de maestros, de la teoría y la práctica del aprendizaje y del diseño de nuevas técnicas instruccionales.

En resumen, una ciencia con algunos rasgos constitutivos claros, pero en la que hay por hacer: una investigación crítica sobre cuál es el modelo explicativo óptimo, la historia de la materia "Psicología de la Educación", concretar y/o limitar o ampliar la notas características del psicólogo de la educación –por ejemplo diferenciarlo del psicólogo escolar–, homogeneizar un programa sobre la enseñanza de la disciplina a nivel de Escuela Normal y de Facultades de Psicología y, finalmente, y sobre todo, una diferenciación, si es posible, con esta otra rama, más moderna y muy cercana a la psicología de la educación, que es la psicología instruccional.

Los metales y la historia

Los metales han estado vinculados al desarrollo del hombre desde la antigüedad con una importancia tan ostensible que incluso la historia está dividida en edades, cuyo nombre coincide con el del metal o con el de la aleación más utilizado en aquel período de tiempo, que, a su vez, corresponde al metal cuya elaboración era más conocida. No es una casualidad que la sucesión histórica sea Edad del Cobre, Edad del Bronce y Edad del Hierro; se trata, sencillamente, de la manifestación de la evolución del arte del beneficio de los metales.

La humanidad se ha desarrollado, hasta hace aproximadamente doscientos años, con sólo ocho metales (cobre, oro, plata, hierro, plomo, estaño, mercurio y cinc) y tres aleaciones (acero, bronce y latón). Sin embargo, actualmente disponemos de setenta metales y de miles de aleaciones. El motivo de esta desproporción en la evolución tecnológica metalúrgica a través del tiempo hay que buscarlo en el hecho de que la metalurgia ha constituido, durante largo tiempo, un arte que se ha convertido, en las últimas veinte décadas, en ciencia y ha dado lugar a una tecnología idónea, capaz de suministrar el material más conveniente para cada aplicación.

El origen de la metalurgia es obscuro debido a la larga distancia cronológica que la separa de nosotros; no obstante, parece ser que el sudoeste del mar Caspio, zona abundante en minerales y en combustibles, fue donde germinó esta tecnología, que después se desarrollaría a otros países limítrofes y lejanos. La aparición de los metales en la escena de los acontecimientos de los pueblos constituye el hito que separa la historia de la prehistoria.

El arte metalúrgico, tal como se ha venido practicando durante los siglos que precedieron a la gran revolución científica del siglo XVIII, fue mera habilidad, cuyas reglas se transmitieron de padres a hijos, de maestros a discípulos. Los descubrimientos, en aquel entonces, eran escasos y aparecían distanciados en el espacio y en el tiempo. Generalmente

se basaban en un accidente, raramente en una acción premeditada.

El hombre primitivo debió emplear, al principio, los pocos metales que encontró en la naturaleza en estado libre, tales como el oro, algo de cobre y acero de origen meteorítico. Hace sólo unos siete mil años que se iniciaron las rudimentarias operaciones de reducción de los óxidos de cobre mediante la combustión del carbón vegetal. El cobre es un metal plástico, pero el descubrimiento de su acritud, o posibilidad de endurecimiento por deformación en frío, intensificó enormemente su aplicación de tal manera que así apareció la Edad del Cobre.

Quizás en un intento de reducción de mena de cobre se calentó, accidentalmente, mineral de estaño y se obtuvo bronce líquido, pues en Armenia ambos minerales aparecen, a menudo, juntos, lo cual nos invita a postular la participación del azar en este descubrimiento. El bronce de estaño satisfizo las necesidades de aquel entonces debido a su resistencia mecánica y química, dureza y colabilidad. Por este motivo la Edad del Bronce sustituye a la del Cobre, hacia unos 3000 años a. C.

El hierro y su aleación con el carbono, el acero, experimentaron cierto retraso en cuanto a su masiva aplicación. El motivo de esta demora no radica en las dificultades de la reducción del mineral de hierro, lo cual se consigue calentándolo con carbón vegetal, sino en las propiedades del producto resultante: masa porosa constituida por mezclas de historia de la materia "Psicología de la Educación", concretar y/ ncluidos en la masa de escoria, sin que hubiera ninguna interacción físico-química entre ambas fases; ello facilitaba la purificación de dicho producto mediante la forja, pues por efecto de la expulsión ("elbowing effect") se desprende la escoria y se sintetizan los glóbulos de hierro. Esta tecnología aparece por primera vez en Mesopotamia hacia el año 1900 a. C.

La obtención del hierro a partir de menas y la fabricación del acero se desarrollaron a ritmo creciente como consecuencia de la abundancia relativa de este metal en la corteza terrestre y, sobre todo, como consecuencia de las propiedades conseguidas mediante la adecuada aplicación de los tratamientos térmicos (temple y revenido) conocidos desde antiguo, principalmente por los hititas. La preponderancia del uso de este metal frente a los descubiertos hasta entonces fue la única causa de la aparición de la Edad del Hierro.

La ingeniería empleada para la extracción del hierro consistió, hasta la aparición del horno alto en Kimberwoth, en el año 1589, en la reducción en fase sólida de los minerales de hierro mediante el carbono y el óxido de carbono obtenido en la combustión del carbón vegetal. Entre los principales procedimientos desarrollados en aquel entonces para esta finalidad destaca el conocido como procedimiento catalán o procedimiento de la farga catalana. Así lo reconocen enciclopedias tan importantes como la "Enciclopedia of Chemical Technology" de Kirk-Othmer. Como sea que actualmente resurgen los procedimientos directos de obtención de hierro, en franca competencia con el alto horno, vuelve a cobrar cierto interés el conocimiento de los empleados antiguamente y, por tanto, el procedimiento de la farga catalana, al que *Investigación y Ciencia* le dedicará un artículo próximamente.

Cromodinámica cuántica

El recurso al concepto de quark ha sido enormemente eficaz para explicar la física de las partículas elementales, pero en qué consista el propio quark no se ha resuelto todavía. De acuerdo con la hipótesis al uso, protones, neutrones y numerosas partículas con ellos relacionadas, los hadrones, no son elementales, sino que están constituidos a partir de otros corpúsculos básicos: los quarks. Así, todos los hadrones conocidos (y superan los 200) pueden explicarse como combinaciones sencillas de algunos quarks. Ha costado más de 15 años delinear una teoría satisfactoria de cómo se disponen los quarks en un hadrón y cómo interactúan entre sí. Aunque la teoría no tiene todos sus cabos atados, hay consenso general en admitir su verosimilitud.

La teoría en cuestión recibe el nombre de cromodinámica cuántica (que se abrevia con las siglas QCD, del inglés "quantum chromodynamics"). Basa su formulación en el modelo de la electrodinámica cuántica (QED, del inglés "quantum electrodynamics"), que es la teoría vigente para explicar las interacciones electromagnéticas. En QED, la magnitud fundamental a describir es la fuerza que surge entre dos partículas dotadas de carga eléctrica, dos electrones por ejemplo. La teoría explica la fuerza al establecer que los electrones pueden intercambiar una partícula intermediaria. Dicha partícula intermediaria es un fotón, partícula carente de masa que se desplaza a la velocidad de la luz. Por su parte, el fotón no porta carga eléctrica sino que transmite la fuerza electromagnética de una partícula a otra.

Las fuerzas que mantienen unidos los quarks no son de tipo electromagnético; antes bien, resultan de otra clase de fuerza de las cuatro que observamos en la naturaleza: la fuerza fuerte. (Las fuerzas restantes son la débil y la gravitacional.) Por lo demás, el mecanismo de interacción entre quarks es muy parecido al que opera en QED. Surge una fuerza entre partículas que tienen una propiedad análoga a la de carga eléctrica, y la fuerza se transmite por otras partículas (que aquí reciben el nombre de gluones porque mantienen unidos a los quarks). Al igual que el fotón, los gluones carecen de masa y se desplazan a la velocidad de la luz.

Una razón de diferencia entre QED y QCD estriba en que los quarks no están condicionados a un solo tipo de carga, sino que pueden portar uno cualquiera de tres. Las cargas reciben el nombre de colores, y de ahí nace la formación de la palabra cromodinámica. En tanto que una partícula eléctricamente cargada sólo puede ser positiva o negativa, una partícula que porte una carga de color puede ser roja o antirroja, azul o antiazul y verde o antiverde. Hay, además, varias partículas intermediarias que se necesitan para transmitir las fuerzas entre los quarks diversamente coloreados. Basta con un fotón para el electromagnetismo, pero se requieren ocho tipos distintos de gluon para una interacción fuerte.

La multiplicidad de cargas de color y de partículas intermediarias contribuyen a que la QCD sea más compleja que la QED, si bien ello no cambia la forma de la teoría. Lo que sí introduce un cambio fundamental es que los propios gluones están asimismo coloreados, a diferencia de lo que ocurre con el fotón, que es eléctricamente neutro. La presencia de una carga cromática en los gluones altera la manera en que la fuerza fuerte varía con la distancia.

La fuerza electromagnética entre dos partículas decrece en relación proporcional con el cuadrado de la distancia que media entre ellas. La fuerza entre quarks es totalmente distinta. Cuando dos quarks se hallan muy próximos entre sí, parecen moverse con cierta libertad e independencia; y, de acuerdo con una hipótesis de trabajo, la intensidad de apareamiento entre ellos llega a anularse cuando la cercanía les permite tocarse. A una escala superior (y en este contexto el diámetro de un protón es una escala grande), el apareamiento se hace mucho más fuerte. Con otros términos, a medida que los quarks se van retrayendo, los enlaces que les mantienen unidos se robustecen. De esta extraña ley de fuerzas se deduce que es difícil, si no imposible, separar un solo quark de un hadrón.

En realidad, no es tan extraña la ley

de fuerzas como parece a primera vista. Incluso en el seno de la QED, la fuerza que ocurre entre partículas no varía de un modo estricto con el inverso del cuadrado de la distancia. Esa relación se cumple sólo a gran escala, y ello resulta en parte del apantallamiento de la carga "desnuda" del electrón al rodear cargas opuestas que aparecen hasta en el vacío. Si se pudiera acercar suficientemente un electrón, la carga eficaz y la fuerza asociada aumentarían. En cromodinámica cuántica observamos justamente el efecto contrario, y ello debido una vez más a que los gluones portan una carga de color. A gran escala, la nube de gluones que rodean un quark tiende a incrementar la carga de color eficaz del quark; cuando dos quarks se encuentran muy próximos, el rápido intercambio de gluones tiende a neutralizar las cargas del quark.

La electrodinámica cuántica constituye la teoría más precisa de la ciencia al haber predicho una constante física con exactitud de una parte en cien millones. La QCD no ha alcanzado tal nivel de refinamiento, si bien en los últimos años se han confirmado algunas de sus predicciones. Así, la teoría sugiere que, si dos hadrones entran en colisión con suficiente violencia, se emitirán un quark y un antiquark en contraposición. Tanto el quark como el antiquark originarán otros hadrones, que formarán "chorros" de partículas que se moverán todas ellas en la misma dirección, aproximadamente. Varios experimentos han detectado la presencia de esos chorros. Y otro experimento, realizado en el acelerador de la Organización Europea para la Investigación Nuclear (CERN), en Ginebra, ha aportado pruebas cuantitativas que corroboran la teoría. El experimento en cuestión empleó neutrinos de alta energía para examinar los quarks dentro de potrones y midió eficazmente la variación de carga de color en relación con la distancia. El resultado de ese estudio europeo concordaba perfectamente con las predicciones de la QCD.

A lo largo de los últimos meses, se han venido desarrollando nuevas técnicas para estimar la probabilidad de determinadas interacciones en el campo de la cromodinámica cuántica. Quizá permitan la predictibilidad de nuevas cantidades y, en consecuencia, lleven a posteriores corroboraciones experimentales de la teoría. En ese proyecto de verificación de determinadas predicciones serán de utilidad nuevos dispositivos experimentales. A mayores energías que las disponibles hoy, una colisión violenta podrá desencadenar tres chorros de partículas, sobre los dos actuales. Esos acontecimientos de triple chorro podrían llegar a observarse en experimentos acometidos en nuevos aceleradores de anillos de almacenamiento a punto de entrar en funcionamiento.

Toda la prueba experimental de la cromodinámica cuántica implica interacciones a baja escala de quarks. Las interacciones a gran escala resultan mucho más difíciles de comprender. La cromodinámica cuántica nos explica, de una manera cualitativa, el enlace fuerte de quarks alejados. Paradójicamente, esa es la razón por la que se ha demostrado inútil a la hora de predecir el movimiento pormenorizado de los quarks a gran escala. Para verificar tales cálculos, QCD y QED siguen el método de sumar una serie de términos; la serie converge rápidamente hacia el valor correcto sólo si el apareamiento entre partículas es débil. Este requisito no suele ser de importancia en electrodinámica cuántica porque la fuerza electromagnética nunca es muy fuerte. El alto valor de la fuerza fuerte restringe el interés de la cromodinámica cuántica a las interacciones de pequeña escala.

Se ha abierto varias vías de solución a este problema. Una de ellas introduce el confinamiento de los quarks como postulado al ubicarlos, atrapados, en un saco cuya superficie no pueden atravesar. De acuerdo con otra hipótesis, los campos de fuerza que mantienen unidos los quarks pueden tener torsiones topológicas, denominadas instantones, que alteran la interacción a gran escala. Una tercera vía que ha enfocado el problema es de orden matemático a través de una técnica propia denominada grupo de renormalización, que permite una solución numérica del problema.

Hay, por último, investigadores que opinan que la incapacidad mostrada por la cromodinámica cuántica a la hora de predecir interacciones a gran escala de quarks no constituye ningún punto débil fundamental de la teoría sino que se trata de una complicación más. Recurriendo a una analogía, el desarrollo de la mecánica cuántica en la década de los años 20 no condujo a la comprensión inmediata de todas las estructuras atómicas, sino tan sólo a la de la estructura más simple: la del hidrógeno. Ello no fue óbice para que se echara mano de la mecánica cuántica y se formularan nuevas teorías sobre el núcleo atómico y, andando el tiempo, sobre las partículas del núcleo. De igual forma, el esfuerzo actual se centra en incorporar la cromodinámica cuántica, a pesar de todas las dificultades, en una teoría superior que podría unificar, cuando menos, tres de las cuatro fuerzas que se registran en la naturaleza: la fuerte, la débil y la electromagnética.

La base molecular del movimiento celular

La capacidad de movimiento no se da tan sólo en las células musculares sino también en otras, y sus diversas formas van desde el deslizamiento celular sobre un sustrato hasta la circulación interna del citoplasma

Elias Lazarides y Jean Paul Revel

El movimiento es una de las propiedades más sorprendentes de la materia viva. Se halla en la base de casi cualquier manifestación de vida y a menudo es la señal más evidente de que algo vive. La motilidad de las células musculares es notable, pero existe un gran número de células no musculares capaces de deslizarse en distintas direcciones y de cambiar activamente su forma. Por ejemplo, los linfocitos y los macrófagos, leucocitos de la sangre que luchan contra las infecciones, son capaces de abandonar el torrente sanguíneo para dirigirse a una herida abierta con el fin de fagocitar las bacterias invasoras. Las células del embrión realizan, durante su desarrollo, una serie de movimientos cuya precisión es digna de un coreógrafo y que lleva a la formación de varias capas de tejido distintas. Por otra parte, las células presentan una gran variedad de movimientos internos como la circulación del citoplasma, la secreción de productos celulares en forma de vesículas, la incorporación de materiales fluidos o sólidos y la separación de cromosomas emparejados en el transcurso de la división celular.

El estudio de los mecanismos de la motilidad celular se ha visto facilitado en los últimos años por el desarrollo de métodos nuevos –y de una forma especial por la utilización de anticuerpos fluorescentes– para localizar proteínas estructurales específicas en el interior de células vivas. Estas investigaciones han llevado a la conclusión de que los diversos tipos de movimiento celular, desde la contracción de fibras musculares hasta la migración de linfocitos, tienen mecanismos moleculares comunes. De la motilidad celular responde un pequeño conjunto de proteínas contráctiles que se hallan presentes tanto en las células musculares como en otras células y cuya función está controlada por un conjunto mayor de proteínas reguladoras. Mediante interacciones sutiles y complejas de estas proteínas entre sí se logra la amplia diversidad de movimientos de que son capaces las células.

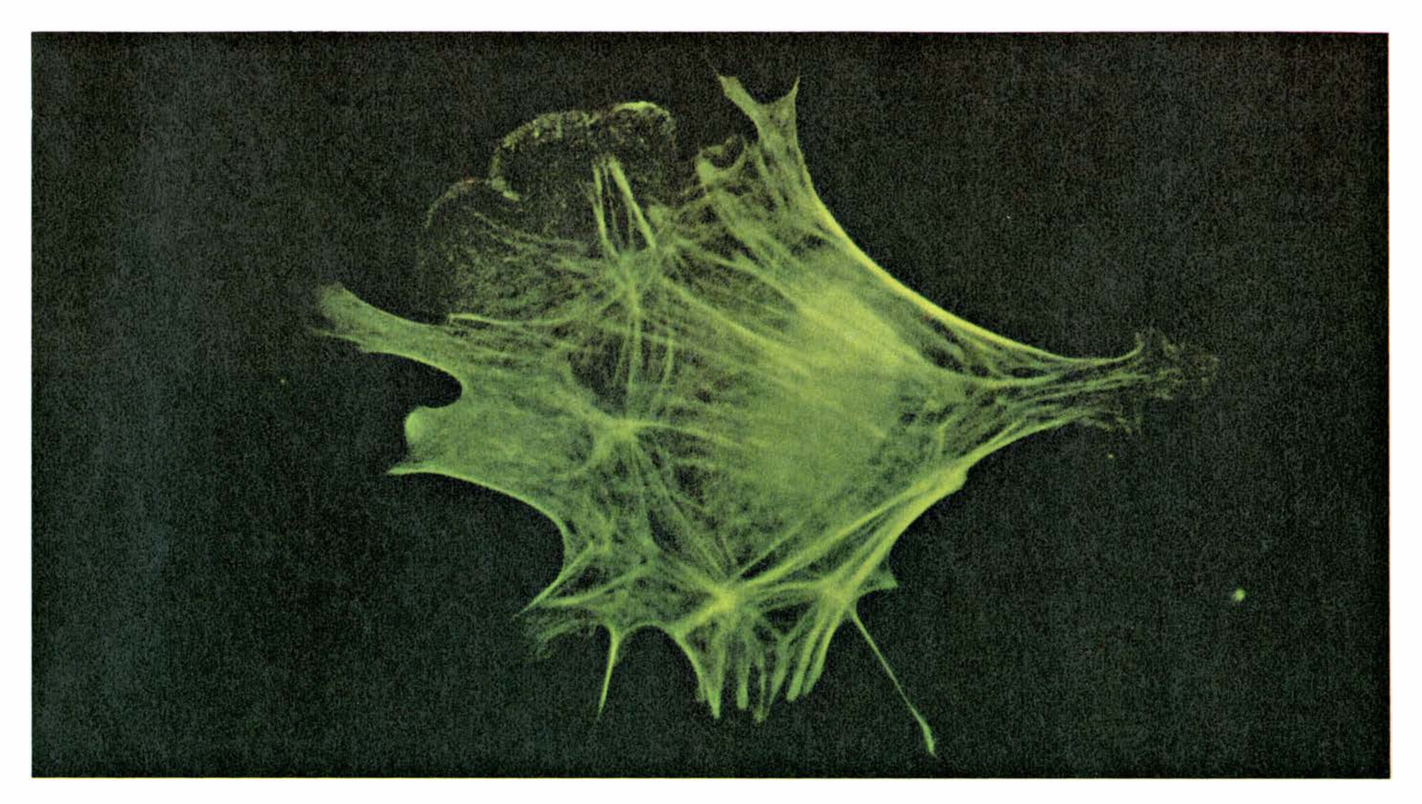
El hecho de que algunas células se mueven se conoce desde que, en el siglo XVII, Anton Van Leeuwenhoek vio nadar espermatozoides en su primitivo microscopio, pero sólo se ha empezado a estudiar seriamente la motilidad celular en este siglo, al desarrollarse métodos para cultivar tejidos animales en el material de vidrio de laboratorio. El trabajo de Warren H. Lewis, del Departamento de Embriología de la Carnegie Institution de Washington, fue de una importancia crucial: para hacer perceptible el movimiento de unas células en cultivo, realizó una serie de fotografías a distintos tiempos. Con el fin de cultivar fibroblastos (células embrionarias de tejido conectivo), trató previamente el tejido con enzimas proteolíticos para romper las zonas de adhesión y las redes de material extracelular que mantienen unidas a las células. Luego separó mecánicamente los diversos grupos de células resultantes y cultivó las células individuales en un medio nutritivo.

Cómo se mueven las células

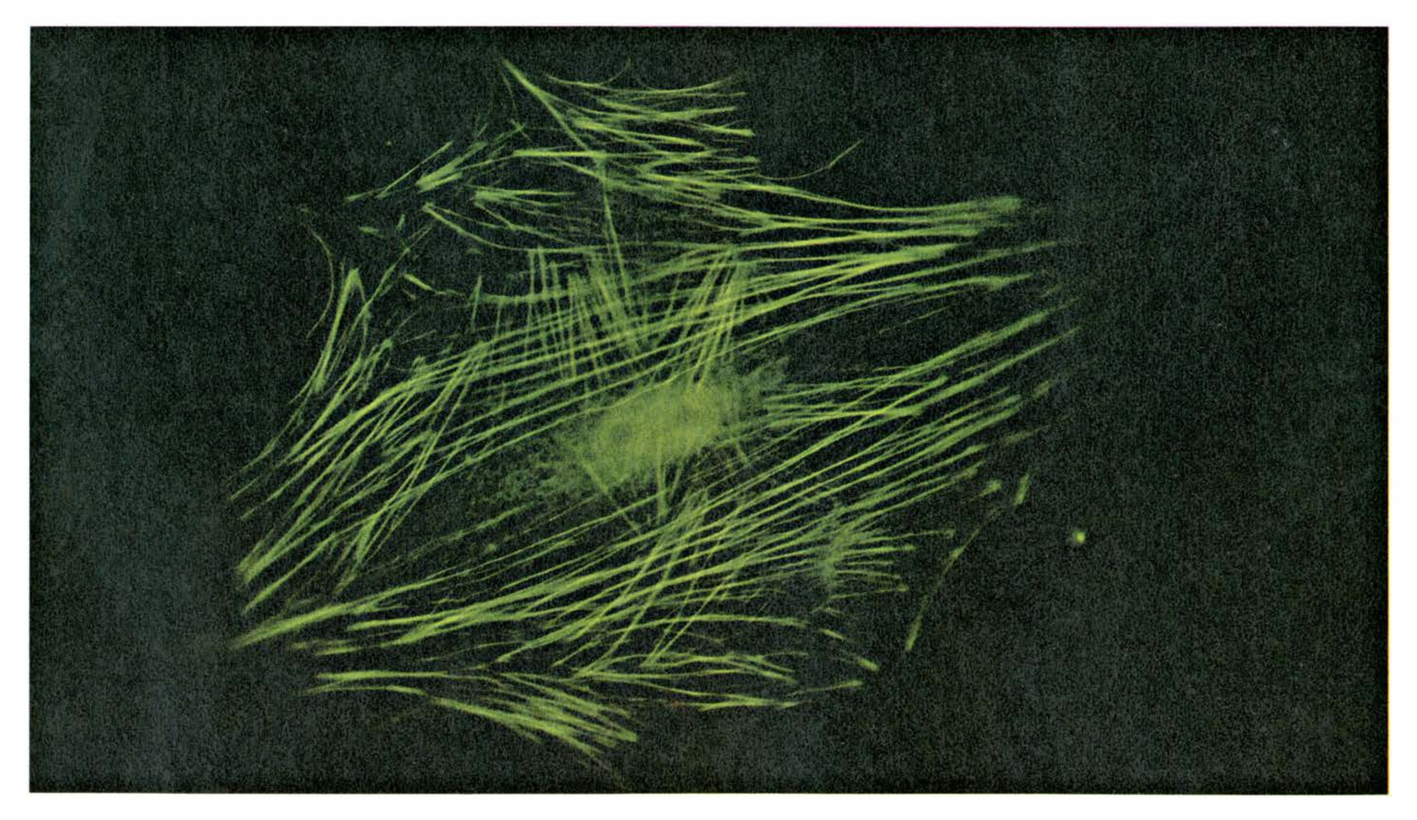
Cuando Lewis examinó al microscopio los fibroblastos que habían permanecido algunas horas en cultivo, éstos habían adquirido una forma poligonal. A lo largo del lado más corto del polígono, el citoplasma formaba unas delicadas prolongaciones laminares denominadas lamelipodios. Los movimientos de los lamelipodios eran tan lentos que resultaban difíciles de discernir, pero la serie temporal de fotografías a intervalos puso en evidencia, según las propias palabras de Lewis, que "los ondulados seudópodos [se desplazan lentamente] hacia delante y detrás como los volantes de un vestido en una brisa ligera". La palabra inglesa "ruffle" –que significa onda o pliegue– es muy evocadora, hasta el punto de que todavía suele emplearse para describir a los lamelipodios y sus movimientos.

Más tarde, Vernon M. Ingram, Michael Abercrombie y colaboradores, del University College London, obtuvieron series temporales de fotografías de fibroblastos en movimiento tomadas desde un lado y no desde encima, con lo que se demostró que la membrana ondulatoria es el órgano locomotor principal de las células. El lamelipodio establece contactos pasajeros con el sustrato sobre el cual la célula descansa, produciendo una extensión del resto de la célula mediante varias adhesiones transitorias. Cuando la célula se ha expandido al máximo, sigue manteniendo uno o dos lamelipodios primarios, tendente cada uno de ellos a conducir la célula en una dirección determinada. Por tanto, la célula migra con frecuentes cambios de dirección, apareciendo éstos de forma típica a intervalos de varias horas.

La secuencia de acontecimientos que representa el movimiento de un fibroblasto a través de una cápsula de cultivo es la que a continuación describimos. En primer lugar el lamelipodio se extiende hacia delante de una forma más o menos paralela al fondo de la cápsula y se pone en contacto con el vidrio. La longitud de esta prolongación laminar suele estar comprendida entre dos y diez micras, pero puede alcanzar veinte micras. Si el extremo del lamelipodio se adhiere al sustrato, la célula tiende a



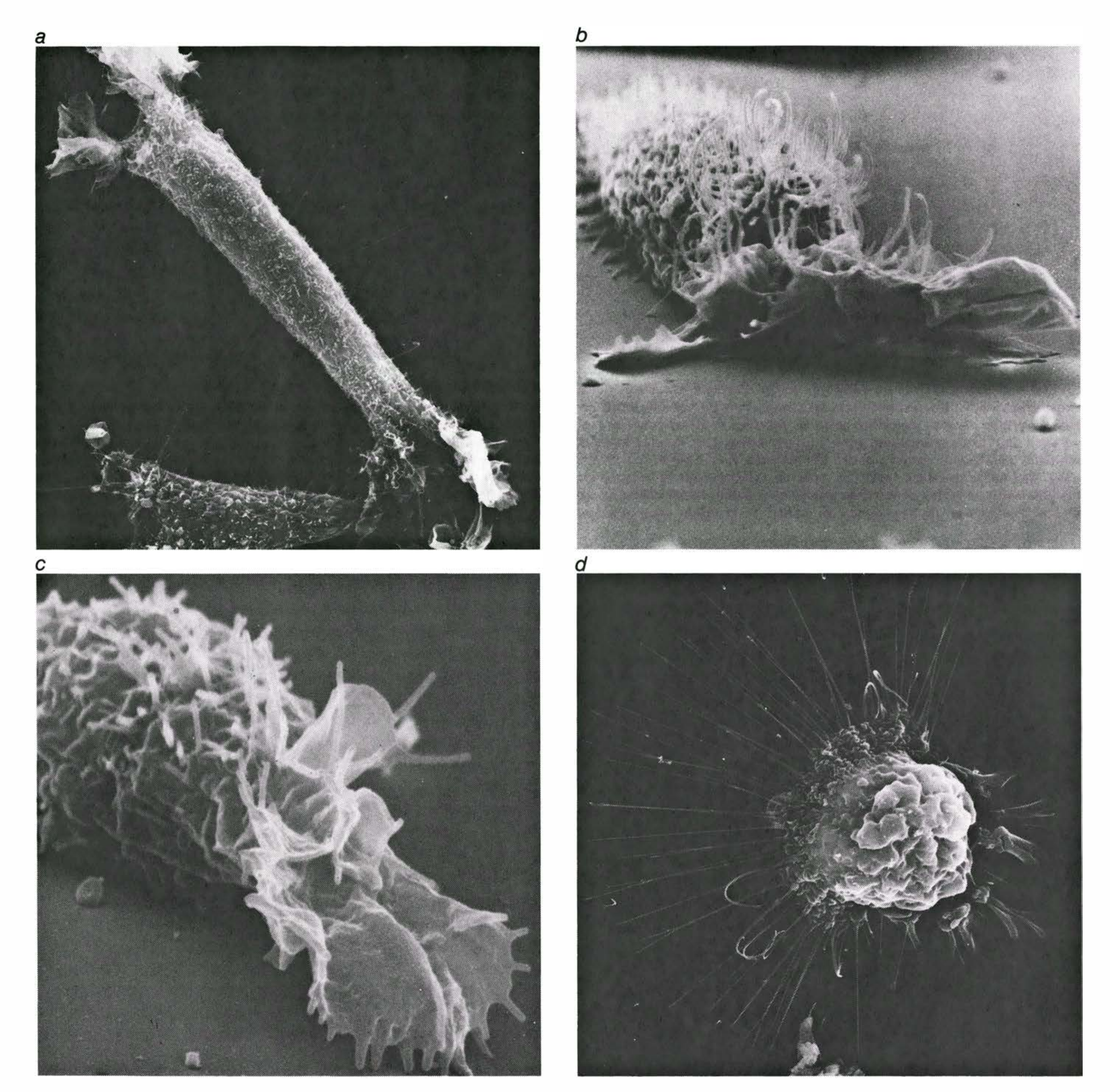
CELULA MIGRANTE de un cultivo de tejido. Presenta una distribución característica de la fluorescencia tras ser tratada con dos tipos de anticuerpos: el primer tipo es específico para la proteína contráctil actina y el segundo tipo tiene especificidad para el primer anticuerpo, estando también marcado mediante un grupo químico fluorescente. Esta técnica, la denominada inmunofluorescencia indirecta, pone en evidencia que la mayoría de los filamentos de actina de la célula migrante se hallan en un estado de red difusa. También se observa que en los bordes de la célula aparecen unas delicadas láminas onduladas del citoplasma que se llaman lamelipodios; éstos forman uniones pasajeras con el sustrato y van arrastrando lentamente a la célula sobre la superficie de la cápsula de cultivo. Esta célula, concretamente, tiene varios lamelipodios y unas cuantas prolongaciones finas, los denominados filopodios, que parecen cumplir una función "sensorial" por detección de otras células migrantes. El núcleo celular, situado en el centro de la célula, presenta una fluorescencia brillante. El aumento de esta micrografía de inmunofluorescencia es de 1500 veces (Elias Lazarides.)



CELULA EN REPOSO extendida sobre la superficie de la cápsula de cultivo. Presenta sus filamentos de actina organizados en fajos lineales, algunos de los cuales van de un extremo a otro de la célula, como se puede visualizar con claridad en esta micrografía de inmunofluorescencia. Cuando la célula empiece a moverse, se disgregarán los fajos de filamentos formando una red difusa semejante a la hallada en la célula migrante de la ilustración superior de esta página. La forma que adoptan los filamentos de actina en el interior celular está directamente correlacionada con su estado de movilidad, de tal modo que, al parecer, la actina desempeña un papel central en el complejo mecanismo molecular del movimiento de la célula.

retraer su margen posterior. Por consiguiente, mientras el extremo frontal se desplaza hacia delante, el resto de la célula se mueve conjuntamente con él. Si el lamelipodio no se pega al sustrato, se repliega sobre la superficie superior de la célula, se colapsa y desaparece fundiéndose con la célula.

La mejor manera de evidenciar los detalles estructurales de la membrana ondulatoria es mediante el microscopio electrónico de barrido, el cual ofrece una mejor resolución y una mayor profundidad de campo que el microscopio óptico. A gran aumento, se comprueba que el lamelipodio aparece como una lámina citoplasmática un tanto irregular, de un espesor aproximado de 0,1 micra, decorada en sus márgenes por pequeños salientes situados a espacios regulares y que recuerdan unos dedos rechonchos. Estos salientes se adhieren a la superficie de la cápsula. También son capaces de prolongarse rápidamente hasta alcanzar varias veces su longitud inicial formando los llamados filopodios o estructuras en forma de bastones. Los filopodios po-



SON CELULAS INDIVIDUALES en cultivo de tejidos lo que queda representado en esta serie de microelectrografías de barrido realizadas por uno de los autores (Revel). En *a* se observa la migración de dos células *LAq* (una línea celular derivada de fibroblastos de ratón) sobre el sustrato. La delicada estructura plumosa de la parte superior izquierda es una "ondulación", o lamelipodio, que caracteriza al extremo frontal de la célula. En *b* se muestra una célula que migra sobre el sustrato, vista desde un ángulo oblicuo, evidenciándose cómo extiende su membrana ondulatoria para formar nuevas zonas de adhesión al sustrato. En la microelectrografía *c* se presenta en primer plano un lamelipodio individual. Las prolongaciones a modo de dedos, colocadas a intervalos regulares a lo largo del borde de la lámina citoplasmática, constituyen lugares de adherencia con el sustrato. En la microelectrografía *d* se observa el aspecto de una célula aplanada que ha sido inducida a redondearse y retraerse mediante incubación durante treinta minutos a cuatro grados centígrados. La célula deja tras de sí largas fibras de retracción que permanecen unidas al sustrato en unas zonas de aplicación amplias y aplanadas que se denominan placas de adhesión. Se puede inducir también el redondeamiento mediante tratamiento con tripsina.

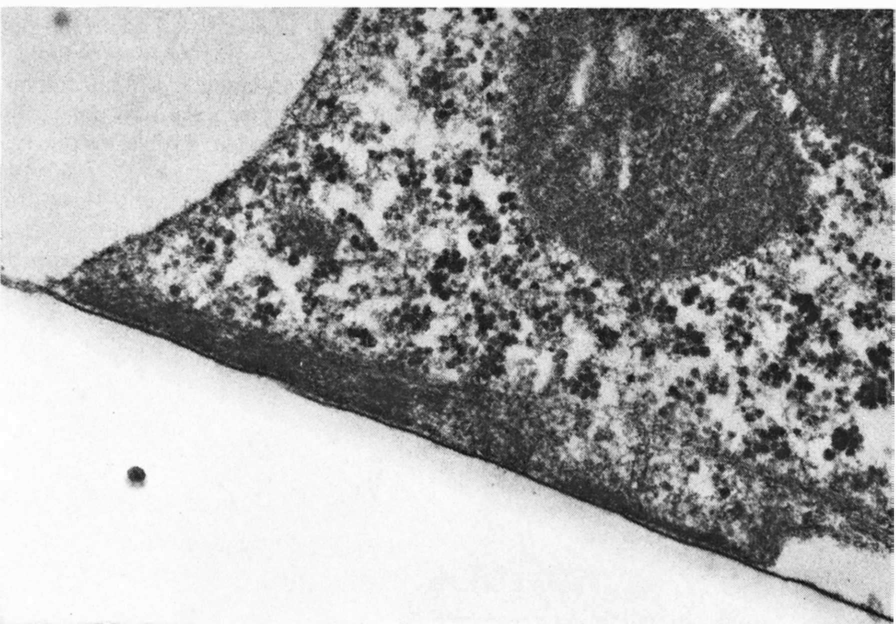
seen un dinamismo elevado, siendo capaces de cambiar del estado fluido al estado sólido en menos de un segundo. Al sobresalir de la superficie celular, sirven como guías espaciales, entremedio de los cuales la membrana celular fluye para formar el lamelipodio. Al constituirse los filopodios, ondean libremente en el medio y por último, o bien se unen al sustrato y se vuelven rígidos, o bien vuelven a fundirse con la célula.

La célula no se adhiere al sustrato con la totalidad de la superficie de la membrana inferior. Lo que hace es establecer contacto en un número reducido de zonas, dejando libre el resto de la membrana, como si se tratara de una mano extendida sobre una mesa cuyo único contacto con la mesa sería a través de las puntas de los dedos. El número de zonas de contacto, denominadas placas de adhesión, puede ser elevado o reducido según el grado de motilidad de la célula: las células más móviles tienen un menor número de placas. Cualquier contacto permanente entre la membrana ondulatoria o un filopodio y el sustrato lleva a la formación de una placa de adhesión y, al migrar la célula sobre el sustrato, las placas se originan y rompen de manera continua. Las placas tienden a quedarse enganchadas al sustrato de modo que, al avanzar la célula, arrastra tras ella largas fibras que van unidas a las placas. Estas fibras de tracción se estiran hasta que son partidas por la tensión y quedan atrás.

Además de su papel en la adhesión y la motilidad celulares, las ondulaciones de la membrana, y más concretamente los filopodios, cumplen una función sensorial. En el desarrollo de un embrión, cuando el filopodio de una neurona que migra se pone en contacto con la superficie de otra neurona, las células cesan el movimiento que las acerca una a otra. Del mismo modo, cuando en un cultivo de tejido el filopodio de una célula que se está expandiendo toca la superficie de otra que ya se ha aplanado, la membrana de la primera célula fluye en dirección opuesta al punto de contacto. Por tanto, parece ser que los movimientos celulares están determinados por algún tipo de computadora química cuya naturaleza está más allá de nuestros conocimientos actuales [véase "La estela de células en movimiento", de Guenter Albrecht-Buehler; INVESTIGACION Y CIENCIA, junio, 1978].

¿Cuáles son los mecanismos moleculares que se encuentran en la base de la motilidad celular? La existencia de este tipo de mecanismo se sugirió en 1952 a raíz de los experimentos realizados por





FAJOS DE FILAMENTOS ACTINICOS se hallan asociados con la membrana en las células animales, hecho que es bien visible en estas dos microelectrografías. La imagen superior se obtuvo fijando a la célula y eliminando posteriormente, mediante un chorro de agua, todo excepto la membrana inferior. Se realizó una réplica del material no eliminado recubriéndolo con partículas de carbono; luego se examinó la réplica al microscopio electrónico. El amplio fragmento de membrana que aparece en la micrografía está anclado al sustrato por medio de un número reducido de placas de adhesión, pero debido a que la membrana está reforzada por los fajos de filamentos estos contactos discretos fueron suficientes para mantener a la membrana adherida. En la micrografía inferior, se observa un corte fino de una célula unida al sustrato y los fajos de filamentos asociados a la membrana inferior de la célula.

Ariel G. Loewy, del Haverford College, con *Physarum polycephalum*, un moho que consiste en una gran masa viscosa de citoplasma. Loewy estaba intentando averiguar cómo un tejido tan poco especializado como éste era capaz de convertir la energía química en trabajo mecánico. Por aquel entonces, tan sólo se estaba empezando a investigar cómo el músculo llevaba a cabo esta misma tarea; no obstante, ya se habían identificado los componentes celulares primarios responsables de la contracción muscular: las proteínas actina y miosina.

Mecanismos de motilidad

Loewy observó que los extractos del moho respondían a la adición de la molécula portadora de energía adenosíntrifosfato (ATP) mediante una disminución de su viscosidad y la eliminación del grupo fosfato terminal portador de energía de la molécula de ATP. Concluyó que estaba analizando un sistema capaz de provocar cambios estructurales bajo la influencia del ATP, es decir, un sistema similar al músculo. La reversibilidad en los cambios de viscosidad le hicieron emitir la hipótesis de que un suministro apropiado de ATP podría tener como efecto que el sistema sufriera alteraciones cíclicas de su estructura y, por tanto, podría provocar la conversión de energía química en el trabajo mecánico que exige la locomoción del moho.

A finales de los años cincuenta, los trabajos de H. E. Huxley y colaboradores, del University College London, dieron a conocer que la contracción del músculo se lleva a cabo mediante el deslizamiento de los filamentos de actina y miosina unos sobre otros. El proceso es dependiente de ATP y está controlado por la presencia o ausencia de iones de calcio. En presencia de calcio, el ATP provoca la formación y disociación cíclicas de puentes entre los filamentos de actina y miosina, de tal modo que los filamentos se van deslizando como por una acción de trinquete.

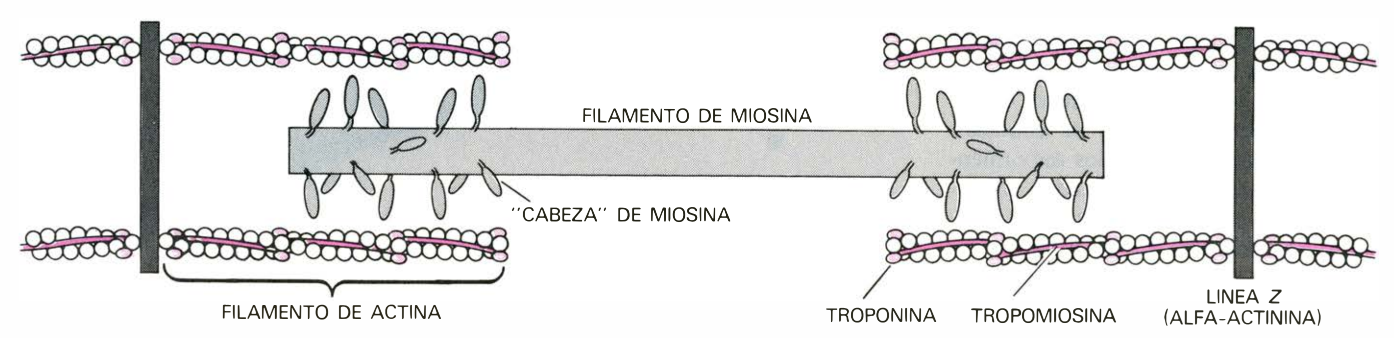
Antes de los años sesenta ya estaba en el aire la idea de que las proteínas contráctiles también podrían estar implicadas en la motilidad de células distintas de las musculares, pero dicha idea todavía estaba lejos de conseguir una amplia aceptación. Uno de los principales obstáculos para la aceptación de la hipótesis era el hecho de que aún no habían sido purificadas la actina y la miosina presentes en células no musculares ni se había comprobado su similitud con la actina y la miosina del músculo. El cambio de puntos de vista se produjo en 1966 cuando Sadashi Hatano y Fumio Osawa, de la Universidad de Nagoya, purificaron actina del moho *Physarum polycephalum*, demostrando que la molécula tiene propiedades físicas, químicas y biológicas muy semejantes a las de la actina muscular.

Actina y miosina de células no musculares

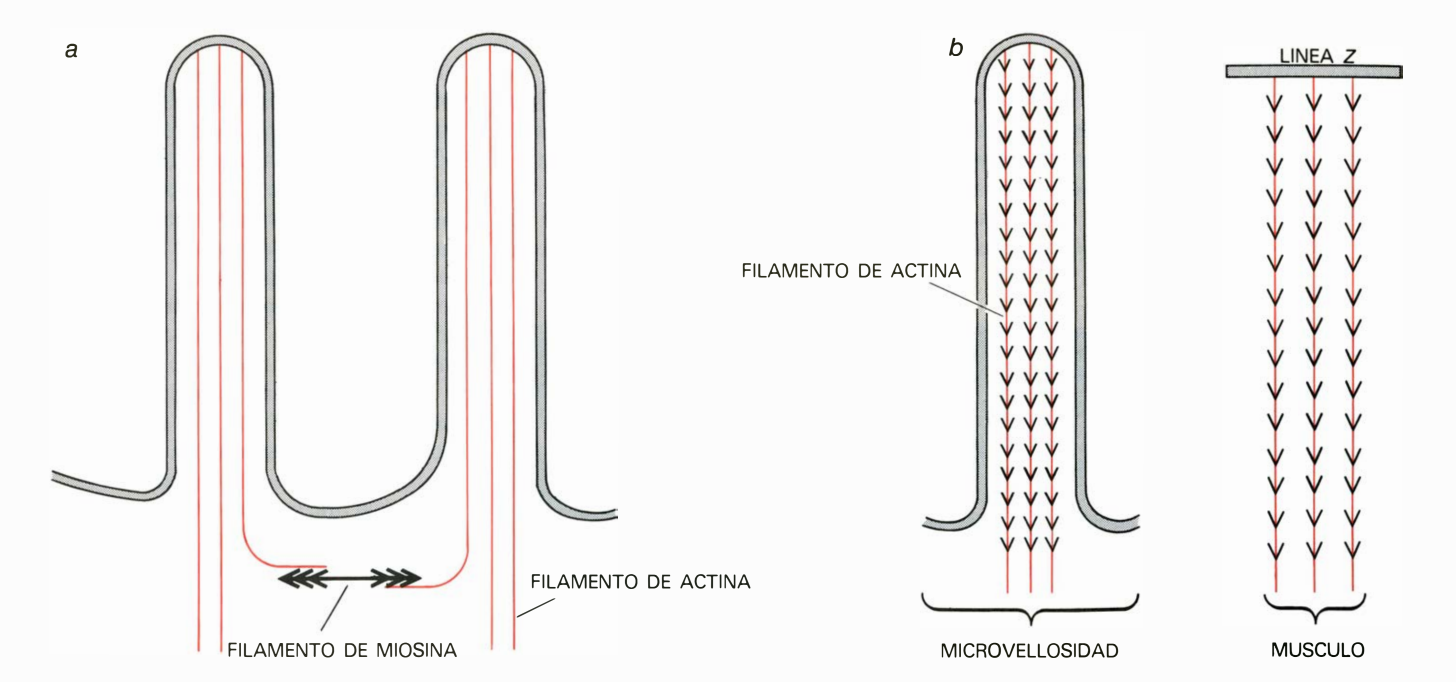
Poco después se identificó la presencia de actina en todo tipo de células no musculares a lo largo de la escala evolutiva, desde los organismos más primitivos hasta los más perfeccionados. La actina llega a representar hasta un 15 por ciento de la proteína total en células que se mueven activamente tales como las amebas y las plaquetas sanguíneas humanas, de un cinco a un 10 por ciento de la proteína en los fibroblastos que migran en un cultivo de tejido, y aproximadamente un dos por ciento de la proteína en tejidos que forman órganos como los que constituyen el hígado de los mamíferos. En el año 1969, Harunori Ishikawa y colaboradores, de la Facultad de Medicina de la Universidad de Pennsylvania, descubrieron que la actina es el principal componente de los microfilamentos, un tipo de fibras intracelulares de singular importancia.

Con el fin de evidenciar la presencia de actina en los microfilamentos, Ishikawa y sus colaboradores utilizaron una técnica que había sido desarrollada por Huxley para estudiar el mecanismo de la contracción muscular. Huxley comprobó que mediante el enzima tripsina era posible separar del resto de la molécula de miosina las "cabezas" que se proyectan fuera de su eje central y que actúan formando puentes cruzados con los filamentos de actina de las células musculares. Observó que los fragmentos correspondientes a las cabezas, denominados meromiosina pesada, conservaban su capacidad para unirse a la actina formando complejos característicos que, examinados al microscopio electrónico, parecen hebras de punta de flecha orientadas a lo largo del eje de los filamentos de actina. Ishikawa y sus colaboradores trataron varios tipos de células no musculares con meromiosina pesada y observaron complejos de puntas de flecha asociados a un tipo de fibras subcelulares, los microfilamentos. Los complejos no estaban asociados a otros tipos de fibras, como los denominados microtúbulos o los filamentos de 100 angstrom. Estos hallazgos indicaban que los diversos tipos de fibras difieren en su composición química. Desde entonces la técnica de la "actina decorada" ha sido muy utilizada como una prueba específica para poner de manifiesto en el interior de las células filamentos que contienen actina.

Los microfilamentos tienen un diámetro aproximado de 60 angstrom y adoptan una serie de configuraciones diferen-



FILAMENTOS INTERDIGITADOS de actina y miosina. Tales filamentos constituyen una estructura muy ordenada en el interior de las células musculares. Aquí presentamos un esquema de un solo sarcómero o complejo de interacción actina-miosina. Los filamentos de actina tienen una polaridad específica y surgen perpendicularmente de las líneas *Z*, o lugares de unión, de modo que los filamentos de lados opuestos tienen polaridad opuesta. Los filamentos de miosina son bipolares, con unas "cabezas" globulares que se proyectan en cada uno de sus extremos; estas cabezas forman puentes cruzados con los filamentos de actina. Existen dos proteínas reguladoras asociadas a la actina, la tropomiosina y la troponina, las cuales permiten la interacción de los filamentos de actina con las cabezas de miosina, pero tan sólo en presencia de iones calcio. La contracción muscular se lleva a cabo mediante el deslizamiento de los filamentos de actina y miosina entre sí, originado por la acción de trinquete de las cabezas de miosina. Parece ser que en las células no musculares se daría una interacción con deslizamiento del tipo existente en el músculo. (Ilustración de Ilil Arbel.)



SE DENOMINA MICROVELLOSIDAD a una minúscula prolongación externa del citoplasma de las células epiteliales que tapizan la cavidad del intestino delgado; sirve para multiplicar varias veces la superficie de absorción de nutrientes. Las microvellosidades contienen filamentos tanto de actina como de miosina y se sabe que se contraen de un modo similar a las fibras musculares; por tanto, constituyen un ejemplo convincente de movimiento no muscular originado por deslizamiento de filamentos de actina y miosina. Como se muestra en *a*, los fajos de filamentos de actina se prolongan en el interior de cada microvellosidad; los filamentos de miosina se sitúan en la base de las microvellosidades. En *b* se determinó la orientación de los filamentos de actina mediante tratamiento de las microvellosidades con fragmentos aislados correspondientes a cabezas de miosina muscular, fragmentos que reciben el nombre de meromiosina pesada; estos fragmentos conservan la capacidad para unirse a los filamentos de actina. Cuando estos fragmentos se aplican a las células musculares, forman con los filamentos de actina complejos de "puntas de flecha" que señalan en la misma dirección que los filamentos. Cuando se añadió la meromiosina pesada a las microvellosidades, las "cabezas" formaron complejos de puntas de flecha con los filamentos de actina dirigidos, desde sus zonas de unión en las puntas de las microvellosidades, hacia el interior celular. Por tanto, los filamentos de actina situados en el interior de las microvellosidades resultan análogos a los conjuntos de ellos hallados en las células musculares.

tes, dependiendo del tipo de célula y del estado en que se encuentra. El examen al microscopio electrónico de células cultivadas permite ver que los microfilamentos recorren una distancia considerable a través del citoplasma, formando como manojos o fajos. Los primeros en identificar estos fajos de microfilamentos fueron Ian Buckley y Keith R. Porter, de la Universidad de Harvard, quienes también sospecharon la importancia de estos fajos en la determinación de la forma ahilada de la célula y en la capacidad celular para adherirse al sustrato y expanderse sobre éste. En una célula migradora individual o en una célula en reposo, la mayoría de los fajos de microfilamentos (que Warren Lewis había propuesto denominar fibras de tensión) se localizan en la zona inmediatamente inferior a la membrana celular y a menudo se hallan en estrecha asociación con ella.

Las placas de adhesión que forman las células en movimiento son precisamente las zonas en que los fajos de microfilamentos establecen contacto con la membrana celular. La mayoría de los fajos están orientados paralelamente al eje longitudinal de la célula y cambian de orientación al cambiar la dirección del movimiento celular. A partir de series fotográficas temporales se ha podido estimar que el período de tiempo que necesitan para reorientarse los fajos varía entre 15 y 60 minutos. En raras ocasiones los fajos de microfilamentos orientados se prolongan hasta alcanzar las partes más externas de la membrana ondulatoria.

Los microfilamentos también pueden presentar formas distintas de la de los fajos rectilíneos. En las células redondeadas que no se adhieren fuertemente al sustrato (como las células en proceso de división o las células tumorales), los microfilamentos forman una red amorfa que difiere notablemente de los fajos de alto nivel de organización. Por lo general, la red se encuentra en la corteza celular, en la proximidad inmediata de la membrana celular (la corteza es una parte del citoplasma que suele estar libre de organelas celulares tales como los ribosomas y las mitocondrias) y, en las células móviles, también es visible en áreas de ondulación de la membrana.

A pesar de que los fajos de microfilamentos aparecen como estructuras de un alto grado de organización, son capaces de disgregarse rápidamente, pasando a la forma de red. Al parecer, los dos estados filamentosos serían ambos extremos de un gradiente y la relación entre ambos vendría determinada por una diversidad de factores como son la actividad de la motilidad celular, el grado de adhesión al sustrato, la región concreta del citoplasma en que se encuentran los filamentos y el estadio de la célula en el ciclo de división. Estas observaciones nos indican que, en una célula procedente de un cultivo de tejidos, los filamentos de actina se organizan en dos sistemas interconvertibles: los fajos, que proporcionan a la célula su fuerza tensora, su capacidad de adhesión y un soporte estructural, y la red, que ofrece además un soporte elástico. Los fajos se observan casi exclusivamente en células inmóviles, mientras que las redes se hallan en las que poseen un movimiento activo. Las células que presentan un movimiento lento contienen una mezcla de ambas formas.

La otra proteína estructural que desempeña un papel en la contracción muscular, la miosina, también pudo ser aislada a partir de células no musculares. Hatano y Masashi Tasawa, de la Universidad de Osaka, fueron los primeros en aislarla del moho *Physarum polycephalum* hacia finales de los años sesenta, y a principios de 1971 Robert S. Adelstein y Thomas D. Pollard purificaron e identificaron la miosina de las pla-

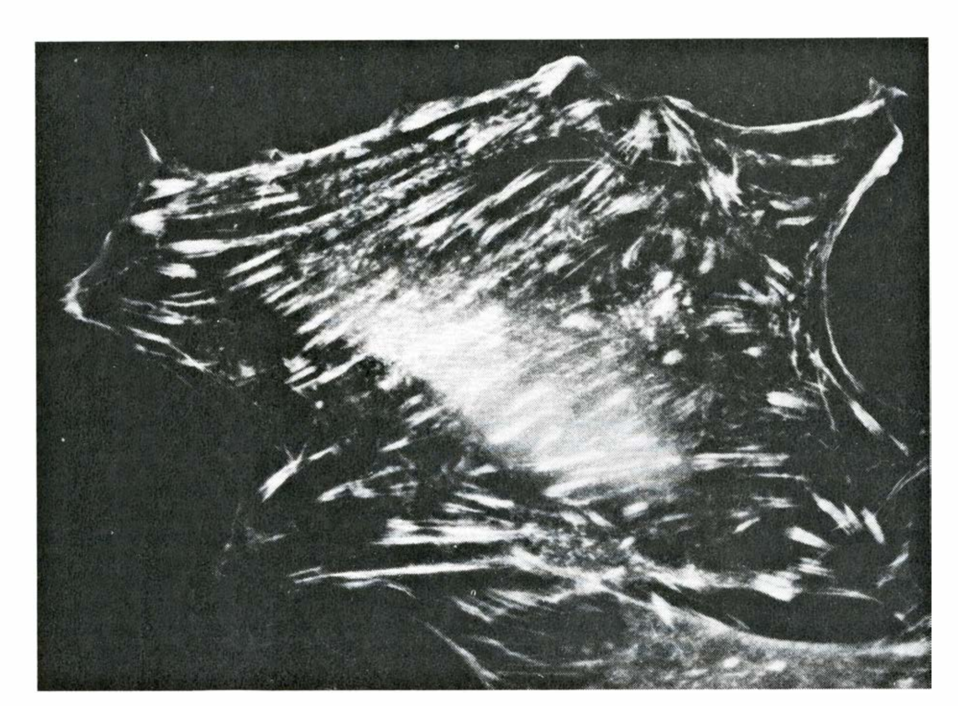
quetas sanguíneas humanas, en su laboratorio del National Institute of Heart and Lung Diseases. Todas las miosinas extraídas de células no musculares comparten la capacidad para hidrolizar al ATP en presencia de actina, propiedad que es característica de la miosina muscular. No obstante, las miosinas no musculares no presentan uniformidad en su capacidad para hidrolizar al ATP en diferentes condiciones iónicas o en presencia o ausencia de actina.

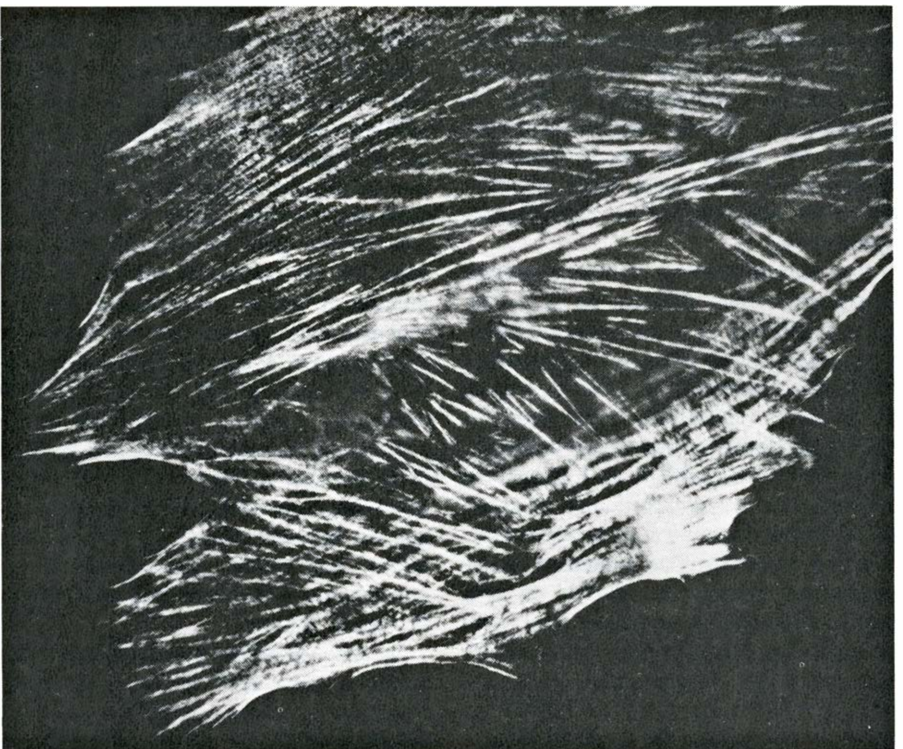
Aunque no se han estudiado con detalle estas diferentes miosinas no musculares, es posible que sean un reflejo de unos papeles funcionales diferentes para estas proteínas. Por ejemplo, la migración de una célula sobre un sustrato se desarrolla durante un período de varios minutos, mientras que la contracción de una fibra muscular se lleva a cabo en cuestión de milisegundos. Es posible, por tanto, que la naturaleza haya evolucionado hacia la constitución de una amplia batería de miosinas, cada una de la cuales tiene una velocidad de consumo de ATP distinta y se regula de un modo diferente según en qué tipo de actividad motora se ve implicada.

Al igual que la actina, la miosina tiene una capacidad inherente para formar filamentos largos. La miosina muscular forma filamentos gruesos cuyo diámetro medio es de 150 angstrom; los filamentos formados por la miosina no muscular son más cortos y más finos, con un diámetro medio de menos de 60 angstrom, es decir, aproximadamente el mismo que el de los filamentos de actina. Así, mientras que la miosina muscular es fácil de identificar en las microelectrografías, los filamentos de miosina no muscular no se distinguen de los filamentos de actina.

La característica que es propia de los filamentos de miosina es que las "cabezas" de sus extremos están orientadas en direcciones opuestas. Por consiguiente, las cabezas pueden formar puentes cruzados con filamentos de actina que presentan direcciones opuestas. Como ya comentamos anteriormente, la contracción muscular se realiza mediante el deslizamiento de los filamentos de actina y miosina unos sobre otros. El modelo de deslizamiento de filamentos fue propuesto por Huxley y se ha adoptado en lo esencial para explicar el papel de la actina y la miosina en la motilidad de las células no musculares.

Una de las diferencias entre las células musculares y las que no lo son es que en las fibras musculares los filamentos de actina están anclados a unas estructuras proteicas aplanadas denominadas líneas *Z* y emplazadas entremedio de cada par de unidades contráctiles. Los filamentos de actina sobresalen perpendicularmente de ambas superficies de la línea *Z* y están orientados en direcciones opuestas a uno y otro lado de la línea. Sin embargo, en las células no musculares los filamentos de actina están anclados directamente a la membrana celular. Por tanto, sería concebible que en las células no musculares las moléculas de miosina fueran capaces de formar puentes cruzados entre dos fajos de filamentos de actina anclados a la parte interna





LAS MICROGRAFIAS DE INMUNOFLUORESCENCIA **hacen visible la distribución de las proteínas contráctiles en el interior de fibroblastos individuales en cultivo. La micrografía superior se realizó con anticuerpos contra la alfa-actinina (una proteína, específica de las líneas *Z* de las células musculares, cuya presencia se ha detectado en células no musculares en las zonas donde los fajos de filamentos actínicos se adhieren a la membrana plasmática). Aparecen manchas de fluorescencia brillantes en dichas zonas de unión, que se distribuyen principalmente por toda la superficie interna de la membrana. La micrografía inferior pone de manifiesto la distribución intracelular de la miosina. Esta segunda toma fue obtenida por Keigi Fujiwara y Thomas D. Pollard, de la Facultad de Harvard.**

de la membrana celular; entonces sería posible que el deslizamiento de los filamentos de actina y miosina acercara más entre sí las dos zonas de unión y, por tanto, tuviera como consecuencia cambios en la forma de la célula.

El gran número de preguntas que permanecían sin contestación acerca de los mecanismos moleculares de la motilidad celular podrían haberse definido con mayor precisión si hubiera habido un modo de seguir la distribución intracelular de las moléculas como la actina y la miosina durante los períodos de actividad motora. El hallazgo de un método de este tipo se convirtió en realidad cuando uno de los autores (Lazarides), que trabajaba entonces en el laboratorio de Cold Spring Harbor en la realización de su tesis doctoral, aplicó al estudio de la motilidad celular una técnica conocida con el nombre de inmunofluorescencia indirecta, desarrollada a principios de los años cincuenta por Albert H. Coons y colaboradores de la Facultad de Medicina de Harvard.

Inmunofluorescencia indirecta

El método de Coons consistía, en primer lugar, en preparar un anticuerpo que se combinara en forma específica con una proteína determinada, la cual intentaba localizar en células humanas cultivadas. Esto lo consiguió inyectando la proteína humana purificada en conejos. Una vez obtenido el anticuerpo de conejo contra la proteína humana, también lo purificaba, fijaba las células humanas cultivadas, las volvía permeables al anticuerpo tratándolas con acetona y luego las incubaba junto con el anticuerpo de conejo. Después de una incubación de una hora, las células eran lavadas y tratadas con un segundo anticuerpo, el cual era obtenido inyectando el anticuerpo de conejo a unas cabras. El anticuerpo de cabra, al cual se había acoplado un colorante fluorescente, se unía al anticuerpo de conejo, el cual a su vez estaba unido a la proteína humana. Mediante el microscopio de fluorescencia, que no sólo sirve para ver las células sino también para someterlas a la radiación ultravioleta, se podía observar el brillo producido por los anticuerpos de cabra que habían formado uniones con la proteína humana y, por tanto, determinar la localización exacta de la proteína en el interior de las células. A la técnica se le denominó inmunofluorescencia indirecta porque empleaba dos tipos de anticuerpo en lugar de uno solo. La ventaja del método consistía en que se lograba una gran amplificación de la fluorescencia, ya que a menudo varias moléculas del anticuerpo de cabra fluorescente se unían a una sola molécula de anticuerpo de conejo.

A pesar de que la inmunofluorescencia indirecta ganó en poco tiempo una aplicación generalizada en todos los campos de la investigación biomédica, durante mucho tiempo no se creyó que pudiera llegar a hacer visible la localización intracelular de la actina y de otras proteínas estructurales. El problema estribaba en que la actina de una especie determinada se parece mucho a la de otras especies y que, por tanto, la actina tiene un bajo poder inductor de formación de anticuerpos. Por ejemplo, cuando se inyecta en conejos actina humana nativa, sólo se fabrican anticuerpos contra las proteínas contaminantes de la actina.

Uno de los autores (Lazarides) resolvió el problema de la siguiente manera. Separó de sus proteínas contaminantes actina procedente de fibroblastos, basándose en las diferencias de peso molecular mediante el uso de una electroforesis sobre el gel en presencia de un detergente fuerte. El detergente provocó la desnaturalización de la actina que adoptó una configuración molecular formando una masa desordenada enrollada al azar. La actina desnaturalizada se eluyó del gel e inyectó en unos conejos para la inducción de anticuerpos.

Sorprendentemente, los anticuerpos contra la actina desnaturalizada reaccionaron no sólo con ella sino también con la actina nativa (no desnaturalizada) y, por otra parte, además de ser activos contra la actina nativa de los fibroblastos, lo era también contra las actinas procedentes de una gran variedad de otros tipos celulares. Al parecer, bastó con la desnaturalización de la actina por el detergente para aumentar hasta un nivel varias veces superior su poder inductor de producción de anticuerpos, aunque los anticuerpos contra el antígeno desnaturalizado conservaron su capacidad para reaccionar con el antígeno nativo. Desde entonces, ya ha sido posible utilizar los anticuerpos antiactina para estudiar la distribución intracelular de la actina mediante inmunofluorescencia indirecta.

Distribución de la actina

Los anticuerpos fluorescentes revelaron la presencia intracelular de un conjunto ordenado de fajos compuestos por filamentos de actina cuya distribución era muy similar a la observada en las microelectrografías. Con frecuencia se hallaban los fajos esparcidos en distintas direcciones y ocupando la totalidad del citoplasma; a menudo también se observaban convergiendo hacia puntos focales y estrechamente asociados con la membrana celular. Su orientación dependía de la actividad motora de la célula: en las células que se movían activamente, se hallaban principalmente paralelos al eje longitudinal de la célula siguiendo la dirección del movimiento, mientras que en las demás células parecían señalar el área nuclear situada en el centro de la célula. La observación más impresionante fue que cada célula tiene su propia ordenación de los fajos de filamentos de actina; en este aspecto, no se pueden encontrar dos células iguales.

Estos resultados tan nuevos consolidaron a la inmunofluorescencia como una excelente herramienta de trabajo para el estudio de las proteínas contráctiles y de otras proteínas estructurales en el interior de las células. La técnica sirve de enlace entre la bioquímica y la morfología de la estructura celular por una parte y las observaciones acerca de la motilidad celular por otra. Aunque sigue siendo la microscopía electrónica la que proporciona el mayor poder resolutivo, la facilidad práctica de la técnica de inmunofluorescencia hace que goce de una mayor popularidad. Cualquier proteína contra la cual se disponga de un anticuerpo es virtualmente localizable en el interior celular, pudiéndose definir su función con todo detalle. Es más, la inmunofluorescencia es el único modo de estudiar la localización de determinadas proteínas que no son identificables al microscopio electrónico: tal es el caso de la miosina no muscular.

La inmunofluorescencia permite la observación intracelular de un gran número de células en diferentes condiciones experimentales y, por tanto, brinda la posibilidad de conocer la función de la actina, de sus proteínas acompañantes y de otros sistemas filamentosos del citoplasma mediante el estudio de su distribución intracelular en el transcurso de distintos tipos de movimiento celular. Ya desde nuestros primeros experimentos podían extraerse conclusiones muy generales. Debido al hecho de que en las células inmóviles la actina se hallaba organizada formando fajos, ganó crédito la idea de que la función primaria de los fajos en este estado celular era de tipo estructural: serviría para mantener a la célula aplanada y expandida sobre el sustrato. En las células muy móviles y con una débil adherencia al sustrato, en cambio, prevalecería el extremo opuesto: en ellas el número de fajos or-

ganizados observable era reducido o nulo y los filamentos de actina se hallaban en su mayoría en un estado que producía tan sólo una fluorescencia difusa, formando como una malla. Esta observación se aplicaba más particularmente a los filamentos de actina situados en las membranas ondulatorias o estrechamente asociados a la membrana celular. No obstante, ello no significa que las células móviles y las inmóviles constituyan dos poblaciones de filamentos de actina mutuamente excluyentes, ya que cualquier población de células cultivadas revela un gradiente entre ambos extremos.

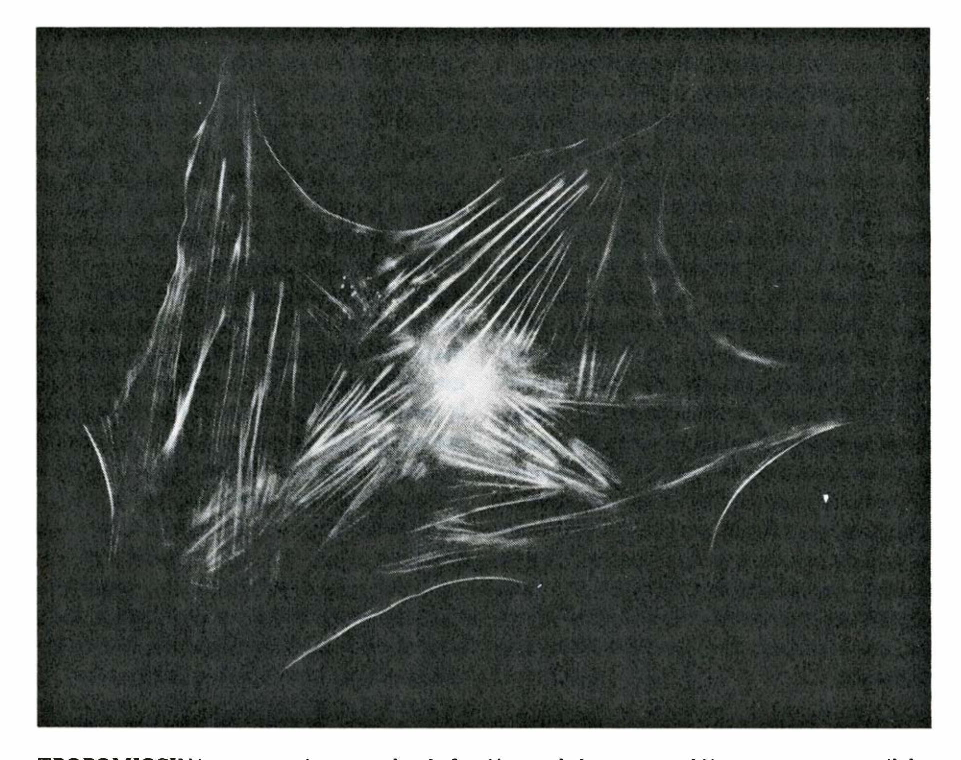
La segunda conclusión que pudimos extraer es que la elevada sensibilidad de la técnica de inmunofluorescencia indirecta permitiría detectar en el interior celular proteínas que se hallaran en cantidades tan diminutas que ningún otro método haría posible su localización. Otro descubrimiento espectacular fue el de que los anticuerpos contra la actina de las células musculares también reaccionan contra las actinas de una gran variedad de tipos celulares, procedentes de especies diferentes. Este hallazgo hacía entrever la posibilidad de fabricar anticuerpos contra otras proteínas estructurales del músculo y de utilizarlos para localizar proteínas relacionadas en células no musculares. Estas esperanzas se confirmaron en los experimentos que describimos a continuación.

Alfa-actinina

El primer aspecto que intentamos esclarecer fue el de las placas de adhesión de las células no musculares: queríamos averiguar si en estas zonas donde los filamentos de actina se unen a la membrana celular, las placas de adhesión son similares, desde el punto de vista bioquímico, a las estructuras de la línea *Z* a las que se hallan unidos los filamentos de actina en las células musculares. También nos interesaba saber si los propios fajos de filamentos de actina llevaban incorporadas estructuras que pudieran funcionar como centros de organización de los filamentos. Para poder comprobar estas hipótesis, necesitábamos un marcador fluorescente que se hallara en las líneas *Z* de las células musculares. El marcador ideal era la alfa-actinina, una proteína conocida como componente de las líneas *Z* (aunque su función sigue siendo un misterio).

Hallamos que los anticuerpos antialfa-actinina muscular reaccionaban con una proteína presente en los fibroblastos y células epiteliales procedentes de cultivos de tejidos. La alfa-actinina se hallaba íntimamente asociada a los filamentos de actina y dicha asociación seguía un orden, hallándose la alfa-actinina no sólo en los fajos de filamentos sino también donde los filamentos estaban relacionados con la membrana celular. La proteína destacaba de un modo especial en las áreas de contacto entre dos células y en las placas de adhesión.

La localización de la alfa-actinina demostraba claramente que las placas de adhesión de las células no musculares son similares a las líneas *Z* de las células musculares. Además, el hecho de que dicha localización se caracterizara por un elevado grado de orden y periodicidad en el interior de los fajos de filamentos de actina demostraba que los fajos estaban organizados, según su eje longitudinal, formando estructuras parecidas a las líneas *Z*. La observación más sorprendente era que, a pesar del elevado grado de organización de los fajos de filamentos, éstos eran capaces de disgregarse en cuestión de segundos. Además de haberse hallado alfa-actinina en las placas de adhesión y en los fajos de filamentos, también se ha descubierto su presencia en las membranas ondulatorias y en los filopodios. Finalmente, llegamos a la conclusión de que cualquier parte de la membrana celular puede funcionar como zona de unión y lugar de organización para los filamentos de actina.

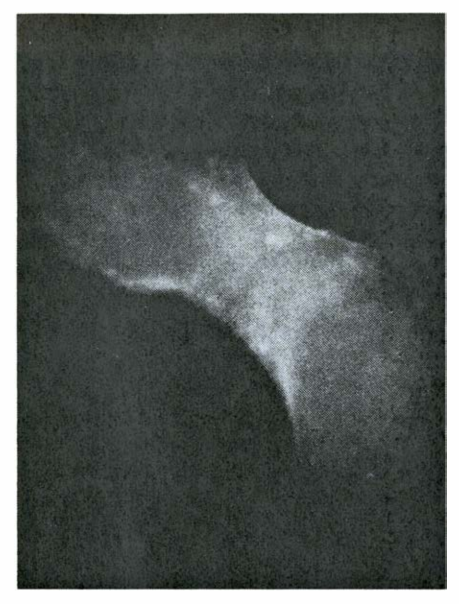
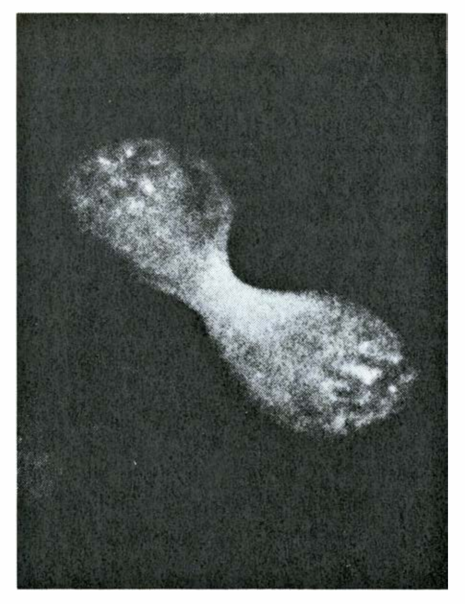
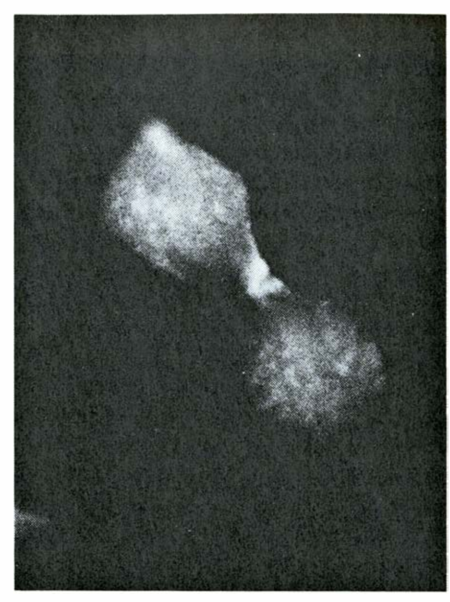


TROPOMIOSINA, una proteína muscular de función reguladora que también se encuentra en células no musculares, hecho que queda patente en esta inmunofluorografía de un fibroblasto cultivado tratado con anticuerpos contra la tropomiosina muscular. La fluorescencia queda uniformemente distribuida a lo largo de los fajos de filamentos de actina. En las fibras musculares, la tropomiosina está encargada de regular la capacidad de los filamentos actínicos para formar puentes cruzados con los filamentos de miosina adyacentes. En las células no musculares, la tropomiosina desempeña un papel diferente: estabiliza los filamentos de actina constituyendo fajos y se halla ausente de la forma de actina difusa.

Orientación de los filamentos de actina

Antes de poder considerar los lugares de unión de los filamentos de actina con la membrana plasmática de células no musculares como estructuras estrictamente análogas a las de las líneas *Z* de las células musculares, era necesario conocer la orientación de los filamentos de actina. Había que saber si señalaban en dirección a los lugares de unión o si se orientaban alejándose de ellos. Richard D. Rodewald y colaboradores, del laboratorio de Morris J. Karnovsky en la Facultad de Medicina de Harvard, como también Mark Mooseker, del laboratorio de Lewis G. Tilney en la Universidad de Pennsylvania, se propusieron hallar respuesta a esta pregunta haciendo reaccionar los filamentos de actina con meromiosina pesada y determinando la orientación de los complejos de puntas de flecha.

Estos investigadores escogieron como sistema modelo de estudio las microvellosidades, o prolongaciones cilíndricas de la membrana celular, del extremo apical "en cepillo" de las células epiteliales

ANILLO CONTRACTIL, formado por fajos de filamentos actínicos y miosínicos, tal como se observa en esta secuencia de micrografías de inmunofluorescencia, procede a dividir una célula madre en dos células hijas. El anillo es una estructura efímera (tiene una vida de unos 10 minutos) que mantiene una anchura y un espesor constantes a la vez que disminuyen de forma continua su circunferencia y su volumen. Por tanto, hay que pensar que algunos de los filamentos actínicos y miosínicos que forman el anillo deben disgregarse tras haber interaccionado entre sí, característica ésta que encuentra difícil explicación en los modelos sencillos de deslizamiento de filamentos. Dado que la mayoría de los tipos de movimiento observados en células no musculares presentan un elevado dinamismo, estos movimientos implican quizá varias interacciones distintas entre filamentos contráctiles.

que tapizan el intestino delgado. El epitelio "en cepillo" se contrae en presencia de iones calcio y ATP, de una forma muy similar a como lo hace la célula muscular, y conserva esta actividad cuando se extirpa del intestino y se baña en una solución de nutrientes. Cada microvellosidad de estas células epiteliales posee de 20 a 30 filamentos de actina; éstos se hallan unidos a la membrana celular en la punta de la microvellosidad y a lo largo de su longitud. En la base de cada microvellosidad, en el lugar donde terminan los filamentos de actina, se hallan presentes filamentos semejantes a los característicos de miosina.

Cuando la microvellosidad se contrae, no varía la longitud de los filamentos de actina, lo que hace pensar con cierta base que el movimiento se consigue por un sistema de deslizamiento de filamentos de actina y miosina parecido al presente en el músculo esquelético.

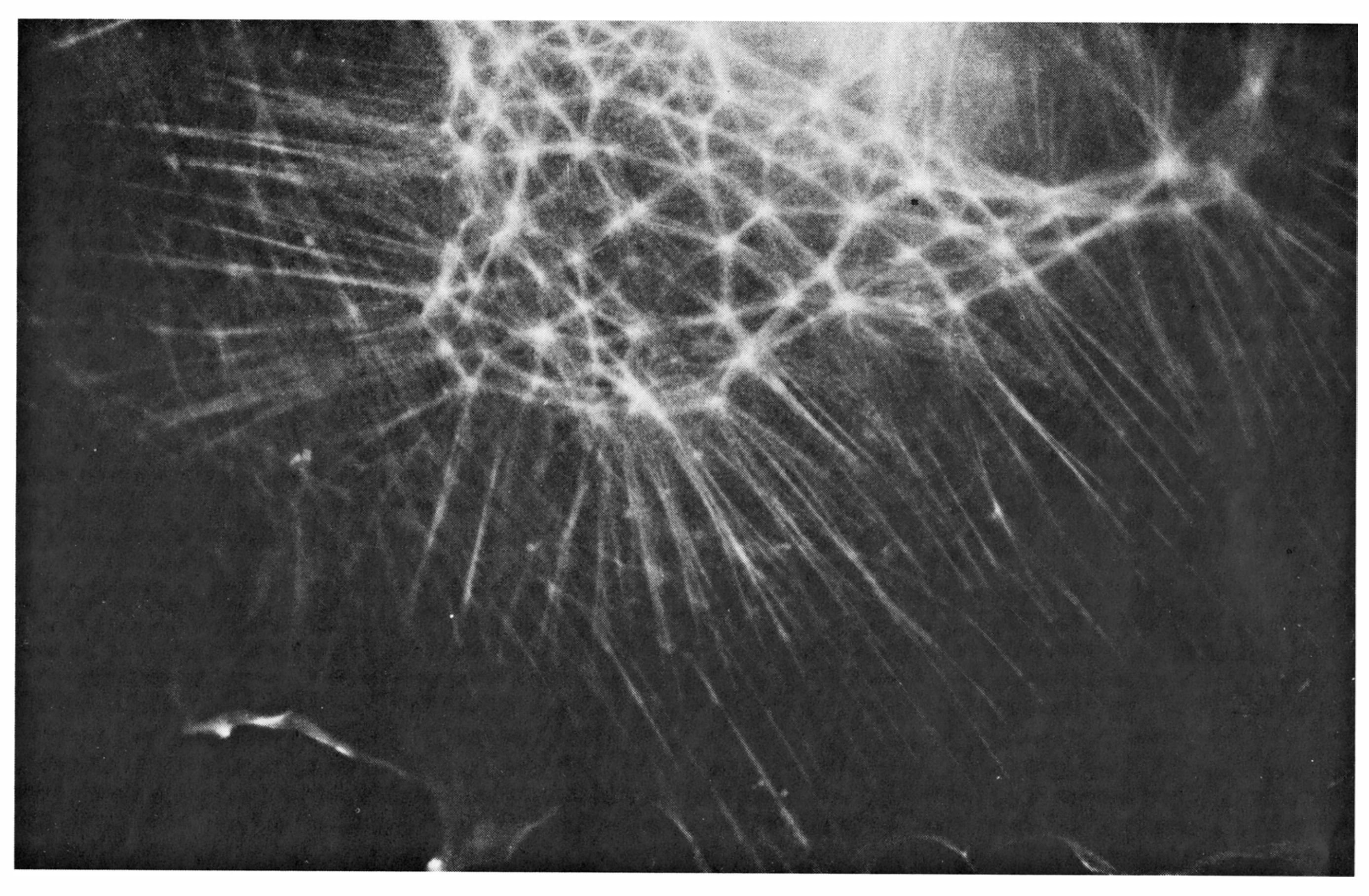
Cuando Mooseker trató a las microvellosidades con pequeños fragmentos correspondientes a "cabezas" de miosina parecidas a la meromiosina pesada, los complejos de puntas de flecha aparecieron a lo largo de los filamentos de actina. Las puntas de flecha señalaban uniformemente hacia abajo, es decir, estaban orientadas desde el punto de unión del filamento situado en la punta de la microvellosidad hacia la base de ésta, indicando una orientación común para todos los microfilamentos. También se han hallado fajos de filamentos de actina unidos a la membrana celular y orientados hacia el interior celular, partiendo de sus puntos de unión, en filopodios de plaquetas sanguíneas y de huevos de erizo de mar. Por tanto, parece ser que la función de algunos, si no todos, los puntos de unión en las células no musculares es análoga a la de las líneas *Z* en el músculo.

El anillo contráctil

Hasta ahora hemos concentrado nuestra atención sobre el papel de los filamentos de actina en el mantenimiento de la forma celular y en los mecanismos que originan el movimiento celular. Aparentemente, la actina también está implicada en el movimiento de partes especializadas del citoplasma. Por ejemplo, durante las etapas finales de la división celular, después de que los cromosomas duplicados hayan sido separados por el huso mitótico, se forma una constricción en el plano ecuatorial de la célula madre, perpendicular al eje longitudinal del huso. Esta constricción ecuatorial o estrangulación celular progresa hasta dividir por la mitad al citoplasma y las organelas, dando lugar a dos células hijas.

En la estrangulación se ejerce una tensión que se puede medir. Esta tensión ha sido estimada en huevos de erizos de mar de distintas especies y alcanza aproximadamente 3×10^5 dinas por centímetro cuadrado, un valor comparable a la tensión desarrollada en el músculo. Thomas E. Schroeder, de la Universidad de Washington, ha demostrado que la tensión ejercida a lo largo de la zona de estrangulación es generada por un anillo contráctil, formado por filamentos de actina alineados constituyendo una circunferencia que ciñe la zona de la constricción celular. Los filamentos de actina están estrechamente asociados a la membrana celular, pero su localización se limita al área de la estrangulación. Keigi Fujiwara y Pollard, de la Facultad de Medicina de Harvard, demostraron poco después la presencia de filamentos de miosina en el anillo contráctil de células humanas en división.

Schroeder estableció la hipótesis de que la constricción del anillo se consigue por interacciones de deslizamiento mutuas entre los filamentos de actina y miosina. La existencia de dichas interacciones pareció confirmarse con los refinados experimentos realizados por Issei Mabuchi y colaboradores, de la Universidad de Tokyo. En lugar de hacer estudios mediante inmunofluorescencia, en los que se requiere que esté fijada la célula entera, inyectaron anticuerpos específicos en células vivas no fijadas y observaron los efectos de este tratamiento sobre la motilidad citoplasmática. Cuando inyectaron anticuerpos antimiosina en embriones de estrella de mar, que se hallaban en el estadio de dos células, en una dosis suficiente para bloquear cualquier interacción actina-miosina en las células, no fue posible la for-



ESTA MALLA REGULAR constituida por fajos de filamentos de actina es un paso intermediario en la conversión de la red difusa de filamentos actínicos, característica de una célula que se mueve activamente, a los fajos altamente organizados que constituyen dichos filamentos en una célula inmóvil, totalmente expandida. Esta malla tridimensional rodea al núcleo celular y sus vértices actúan como centros organizadores para los fajos de filamentos actínicos que se prolongan hacia la zona periférica de la célula. Constituidos estos fajos periféricos, la propia malla es convertida a la forma de fajos lineales. Debido al sorprendente parecido de esta malla con las cúpulas de Fuller, los autores le han dado nombre de geodomo celular.

mación de una nueva estrangulación y el huevo dejó de dividirse. Estos resultados demuestran claramente que es necesaria una interacción entre la actina y la miosina para que pueda tener lugar la segmentación celular.

El anillo contráctil es una estructura pasajera que tan sólo persiste unos 10 minutos; la rapidez con que se constituye y se disgrega da fe de su naturaleza extraordinariamente dinámica. Schroeder comprobó que, a pesar de que la anchura y el espesor del anillo permanecen constantes durante su constricción, su volumen disminuye. Por tanto, tiene que haber una disgregación parcial del anillo durante su contracción, es decir: cualquier interacción de deslizamiento entre los filamentos de actina y miosina debe ir seguida por la desorganización de algunos de los filamentos. No obstante, el hecho de que la anchura y el espesor del anillo permanezcan constantes durante la estrangulación significa que dicha disgregación progresiva se debe producir uniformemente en todo el anillo durante todo el período de su efímera vida. Por ello, y a pesar de la presencia de actina y miosina en el anillo contráctil, un modelo sencillo basado en el deslizamiento de los filamentos resulta insuficiente para explicar la segmentación celular.

Aunque todavía son excesivamente grandes nuestras lagunas en torno a su base molecular para que pueda ser formulado un modelo detallado del anillo contráctil, estos hallazgos subrayan el hecho de que la mayoría de los acontecimientos de motilidad que se dan en las células no musculares son dinámicos y pasajeros y no son, por tanto, estrictamente análogos a la contracción muscular. Además, la simple asociación de la presencia de actina y miosina con un acontecimiento de motilidad determinado no significa que estas moléculas desempeñen en él un papel funcional. Por ejemplo, se ha demostrado que tanto la actina como la miosina son componentes del huso mitótico, el cual está constituido principalmente por microtúbulos. No obstante, unos estudios basados en microinyecciones de anticuerpos han demostrado que probablemente no sea necesaria una interacción actina-miosina para el movimiento de los cromosomas.

Proteínas reguladoras de la actina

La siguiente cuestión que se planteó fue el saber si alguna de los proteínas implicadas en la regulación de la contracción muscular podía darse en asociación con la actina de células no musculares. En el músculo, la interacción entre los filamentos de actina y miosina está regulada por dos proteínas: la tropomiosina y la troponina. Cuando son bajos los niveles de calcio en el interior de la célula muscular, un complejo constituido por ambas proteínas inhibe la interacción actina-miosina. En cambio, cuando los niveles de calcio son elevados, éste se une a la troponina, la cual a su vez actúa sobre la tropomiosina, iniciando una secuencia de cambios conformacionales que permiten la interacción de la actina y la miosina.

La tropomiosina es una de las moléculas reguladoras más universales: se la encuentra asociada a los filamentos de actina musculares en una amplia escala

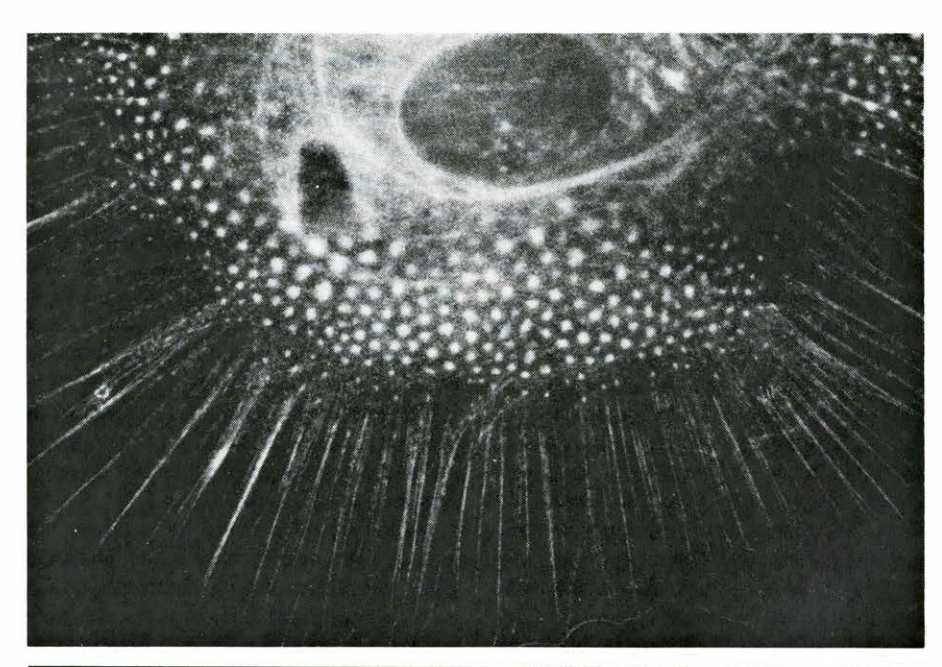
de organismos que va desde los moluscos hasta el hombre. La inmunofluorescencia nos proporcionó entonces una oportunidad única para averiguar si los filamentos de actina observados en las células no musculares también llevaban tropomiosina asociada. Pensábamos que la presencia o ausencia de tropomiosina podría ser un reflejo de importantes diferencias en cuanto a regulación entre conjuntos de filamentos de actina implicados en distintas actividades celulares.

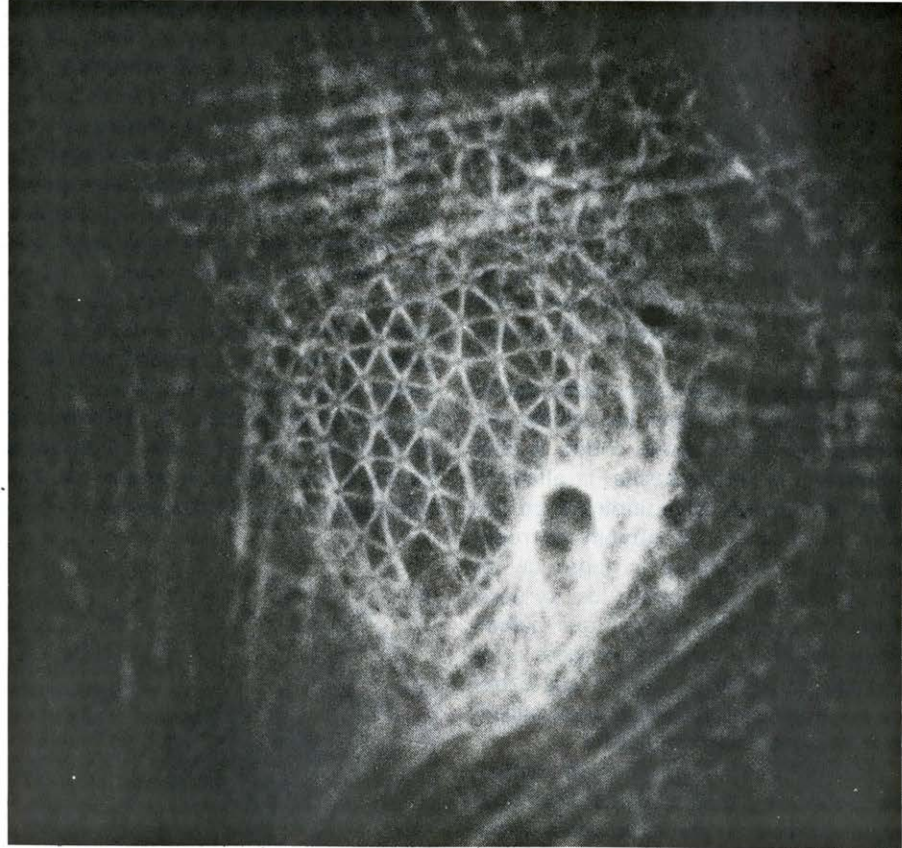
Cuál no fue nuestro asombro al comprobar que en las membranas ondulatorias y los filopodios activos era débil, o nula, la fluorescencia debida a la tropomiosina y que tampoco presentaban gran cantidad de ella los fajos de filamentos de actina pertenecientes a células con gran actividad motora. Aunque la ausencia de fluorescencia no implica necesariamente que la tropomiosina esté ausente de estas áreas (es posible que contengan una tropomiosina de distinta actividad inmunológica que la tropomiosina muscular), existen otras observaciones que parecen confirmar esta conclusión. En efecto, nosotros observamos que, cuando la célula está fuertemente adherida o inmóvil, sí se encuentra tropomiosina a lo largo de los fajos de filamentos de actina y en las membranas ondulatorias.

Todo ello parece indicar que el papel de la tropomiosina en las células no musculares es diferente del que desempeña en el músculo. En las células no musculares, la proteína se une a filamentos de actina que participan en el mantenimiento de la forma celular y no en las actividades relacionadas con la motilidad. En consecuencia, la presencia de tropomiosina refleja la presencia de filamentos de actina con funciones estructurales y su ausencia refleja la presencia de filamentos de actina con funciones motoras. Así pues, algunos de los filamentos de actina de las células no musculares no forman uniones con la tropomiosina muscular, actuando de distinta manera, por tanto, que los filamentos de actina de las células musculares.

Puesto que las células no musculares contienen tropomiosina, por analogía se podría esperar que también contuvieran troponina. De momento, las pruebas de la existencia de troponina en células no musculares son poco concluyentes; no obstante, se han identificado proteínas análogas a la subunidad de la troponina muscular que se une al calcio, denominada troponina *C*. Estas observaciones indican que, al igual que ocurre en el músculo, los iones calcio son reguladores importantes de la motilidad en las células no musculares. De hecho, parece que el calcio controla el paso del estado inactivo al activo –o viceversa– de muchas moléculas que participan de un modo u otro en la motilidad celular.

A fin de poder estudiar con mayor detalle la asociación con los filamentos de actina de moléculas tales como la alfa-actinina y la tropomiosina, intentamos buscar un sistema experimental que nos permitiera seguir la conversión de una célula de un estado de gran movilidad a otro inmóvil, con adherencia al





LAS PROTEINAS ACOMPAÑANTES que participan en la construcción del geodomo celular han sido puestas en evidencia en estas dos inmunofluorografías, una realizada con anticuerpos antialfa-actinina (*arriba*) y la otra con anticuerpos antitropomiosina (*abajo*). Las imágenes indican que la malla del geodomo consiste en unos puntos focales que contienen actina y alfa-actinina, focos que están conectados entre sí por filamentos de actina asociada a tropomiosina. Los filamentos actínicos lineales que surgen de los vértices de la malla del geodomo contienen tanto tropomiosina como alfa-actinina.

sustrato. Se obtuvo un sistema de estas características al tratar las células con bajas concentraciones de un enzima proteolítico como la tripsina. La célula tratada enzimáticamente retrae su citoplasma y su membrana celular acercándolos al área nuclear y su adhesión al sustrato queda reducida tan sólo a las fibras de retracción. En el proceso de cambio de forma celular, de la disposición aplanada a la redondeada, los fajos de filamentos de actina se convierten rápidamente en una red filamentosa.

Cuando se deja que las células despegadas vuelvan a instalarse sobre el sustrato, sufren un aplanamiento lento que cubre un período de seis a ocho horas y va acompañado de una intensa actividad ondulatoria en la zona periférica. Durante este tiempo, la actina se encuentra casi exclusivamente en la forma de red y se localiza principalmente en la áreas de ondulación de la membrana. Sin embargo, poco a poco y de una forma gradual, los filamentos de actina se reorganizan formando fajos, con una reducción concomitante de la actividad ondulatoria de los bordes celulares. Este proceso posibilita el estudio de la constitución de los fajos.

Geodomo celular

Observamos que muchas células organizaban sus filamentos de actina para formar los fajos de modo que éstos se distribuían en una red sorprendentemente regular, que se asemejaba en gran manera a una cúpula geodésica. La red ocupaba la totalidad del área que se halla por encima y alrededor del propio núcleo.

Asimismo se observaron redes regulares semejantes al tratar las células con anticuerpos fluorescentes contra la alfa-actinina y la tropomiosina: la alfa-actinina se localiza esencialmente en los vértices de la red; en cambio, la tropomiosina se hallaba a lo largo de las cortas fibras que unían los vértices. También pudimos destacar que habían fibras de mayor longitud ancladas a los vértices de la red en uno de sus extremos y que se prolongaban hasta alcanzar la membrana ondulatoria.

Estas últimas fibras contenían actina, alfa-actinina y tropomiosina ordenadas según un patrón que no se podía distinguir del hallado en los fajos de filamentos actínicos en las células totalmente expandidas.

También se observaba con relativa frecuencia que los fajos de filamentos de actina, originados en los vértices de la red en forma de cúpula geodésica, se prolongaban en el interior de los filopodios o lamelipodios. Por tanto, los vértices de la red actúan como centros de organización para los fajos de filamentos actínicos responsables de mantenimiento de la estructura de estas prolongaciones celulares. Cuando decidimos seguir la formación de la estructura en red mediante inmunofluorescencia indirecta, comprobamos que tanto la actina como la alfa-actinina se reunían en uno de los vértices en los primeros estadios de la expansión celular. Luego, transcurría un lapso de tiempo considerable antes de que la tropomiosina se pudiera detectar en asociación con la red, lo cual hace suponer que la alfa-actinina es la que determina los lugares de unión de la tropomiosina sobre los fajos de filamentos de actina. Estas observaciones también aportaron pruebas adicionales de que los filamentos de actina presentes en la célula no necesariamente han de tener tropomiosina asociada.

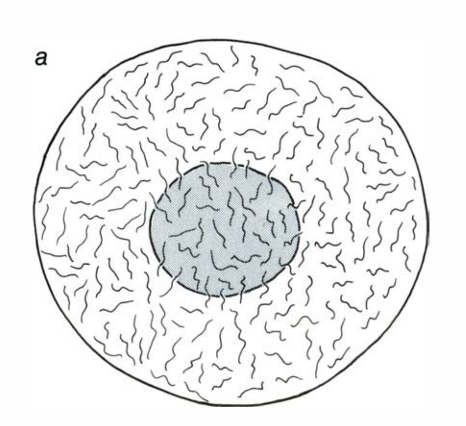
Poco tiempo después de la aplicación de la inmunofluorescencia indirecta para la localización de la actina, la técnica fue adoptada por otros investigadores para la localización en el citoplasma de otras proteínas contráctiles y estructurales, como las componentes de los microtúbulos y de los filamentos de 100 angstrom. Varios colaboradores han logrado demostrar que la miosina se localiza a lo largo de los fajos de filamentos actínicos, siendo su distribución variable según la actividad motora de la célula. La existencia de una estrecha asociación entre tanto la miosina como la tropomiosina y los fajos de filamentos de actina parece confirmar la idea de que estos filamentos son responsables de algunos fenómenos de motilidad originados por un mecanismo de deslizamiento de filamentos. No obstante, no se sabe si los fajos son realmente capaces de contraerse por dicho mecanismo.

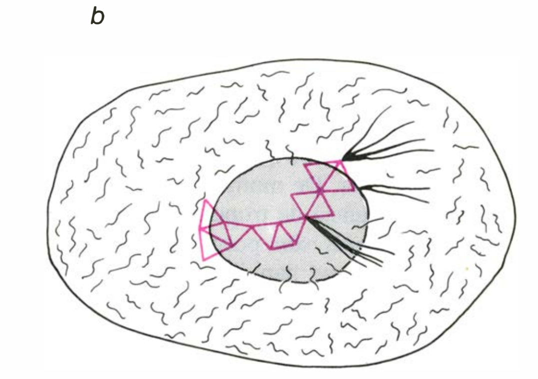
Una célula que se halla en un cultivo de tejido puede estar en cualquier punto situado entre estos dos extremos: el de descanso total y el de acentuada movilidad. Las diferentes partes de la célula como los filopodios y las membranas ondulatorias también presentan un grado de motilidad variable. En las células que se mueven con lentitud, hay partes del citoplasma que están inmóviles mientras que los filamentos de actina van oscilando, rápida o lentamente, entre el estado de fajos y el de red. Pero ¿cuáles son los factores citoplasmáticos que regulan la transición de los filamentos de actina entre ambos estados? ¿Cómo se mantiene la estructura celular cuando los fajos de filamentos actínicos se han transformado en una red? ¿Cómo es posible que la red sea inductora de movimiento?

La regulación de la actina

Algunas de estas preguntas hallaron respuesta de forma inesperada a raíz de unos experimentos realizados *in vitro* relativos a la polimerización de la actina. Fue Robert E. Kane, de la Universidad de Hawai, quien al estudiar dicha polimerización observó que los extractos citoplasmáticos de huevos de erizos de mar formaban un gel de modo espontáneo al elevarles la temperatura de cuatro grados a 37 grados centígrados y que este gel era rico en filamentos de actina polimerizada.

Posteriormente, Kane procedió a demostrar que la gelación venía inducida por la conversión de la actina monomérica (no polimerizada) en filamentos de actina. Para hacer posible la gelación, se requería también la presencia de dos proteínas adicionales: una proteína cuyo peso molecular es de 58.000 dalton y que actúa sobre los filamentos de actina individuales ordenándolos en fajos, y





CONVERSION DEL GEL EN FAJOS, en una célula cultivada que se está aplanando. Dicha conversión exige la organización transitoria de los filamentos de actina en una estructura regular de geodomo. Posteriormente, los vértices de esta malla actúan de centros organizadores para los fajos de filamentos

otra, de un peso molecular de 220.000 dalton, que forma uniones cruzadas entre los fajos para construir el gel tridimensional.

Poco después, Thomas P. Stossel, del Children's Hospital Medical Center, descubrió que los extractos citoplasmáticos de macrófagos de pulmón de conejo también formaban un gel. Sin embargo, en este caso tan sólo se requería la presencia de los filamentos de actina y de la proteína que se une a la actina (de 220.000 dalton) para conseguir la formación de un gel tridimensional constituido por uniones cruzadas. Al microscopio electrónico, el gel de actina tenía una apariencia muy similar a la red de filamentos de actina descrita por otros autores en las áreas de elevada motilidad de la célula, como pueden ser la membrana ondulatoria y los filopodios. Por otra parte, el gel de actina se contraía de forma reversible en presencia de miosina. Pollard y D. Lansing Taylor obtuvieron resultados similares con actina aislada de amebas.

Estos resultados dieron una nueva dimensión al conocimiento de la base molecular de la motilidad de las células. Es posible que los geles de actina sean responsables de proporcionar un sostén mecánico a las estructuras citoplasmáticas dinámicas como la membrana ondulatoria, los filopodios y la corteza celular. Así, la motilidad podría ser consecuencia de contracciones reversibles del gel de actina, inducidas por la miosina. El citoplasma de las células que se caracterizan por una elevada motilidad, como los macrófagos, estaría atravesado por la trama del gel de actina, siendo así el citoplasma una masa viscosa que serviría de sostén elástico para las actividades motoras de la célula.

Otros investigadores no tardaron en identificar, en varios tipos de células musculares, proteínas parecidas a la proteína de los macrófagos que se une a la actina. Kuan Wang, del laboratorio de Jon Singer de la Universidad de California en San Diego, aisló de células de musculatura lisa una proteína denominada filamina cuyas propiedades eran muy similares a las de la proteína que se une a la actina. Por estudios de inmunofluorescencia indirecta, se comprobó que en las células no musculares la filamina se halla situada de una forma periódica a lo largo de los fajos de filamentos actínicos pero se encuentra en una forma difusa en la membrana ondulatoria y en la corteza celular. Se ha podido demostrar que "in vitro" la filamina induce la agregación de la actina para formar un gel filamentoso.

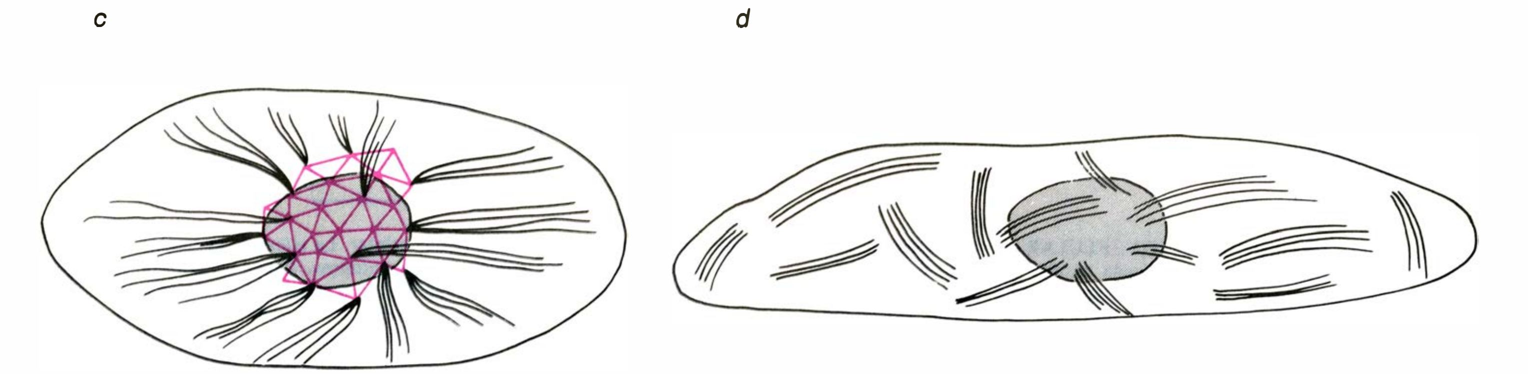
Control de la polimerización de la actina

No siempre la actina se halla en una de las dos formas filamentosas (de fajos o de red): también se hallan presentes en el citoplasma cantidades considerables de actina monomérica. Unas primeras observaciones indicativas de cómo se logra la regulación de los niveles de actina no polimerizada surgieron de los trabajos de Uno Lindberg y colaboradores, de la Universidad de Uppsala. En 1974, lograron la purificación y ulterior cristalización de actina monomérica de bazo de ternero. El análisis estructural de los cristales de actina reveló que la actina monomérica se hallaba asociada con cantidades equivalentes de una proteína de bajo peso molecular. Los experimentos posteriores demostraron que esta proteína se responsabiliza de mantener a la actina en su estado monomérico y que, al eliminarla, se permite que la actina se polimerice formando filamentos. Lindberg dio a esta proteína el nombre de profilina. El descubrimiento de esta proteína reguladora ha ampliado todavía más el campo que permitirá entender la motilidad celular a nivel molecular.

Las observaciones que hasta ahora hemos estado comentando nos indican la gran diversidad de las vías de regulación de la actina. Se han intensificado proteínas que regulan (1) la transformación de actina monomérica en filamentos poliméricos, (2) la agregación de estos filamentos para formar fajos o un gel con estructura de red, (3) la transformación de los fajos en la forma de red y (4) la asociación de los filamentos de actina con la membrana celular. Además de estos controles proteicos, hay que tener en cuenta que existen actinas químicamente diferentes que se polimerizan en condiciones iónicas diferentes, como puede ser en presencia o en ausencia de iones calcio. Es posible que la célula también imponga otros tipos de control como variaciones de *p*H o activación y desactivación de las proteínas reguladoras mediante hidrólisis enzimática o modificaciones químicas.

La regulación de la miosina

Ya que la miosina participa en algunos, si no todos, los tipos de motilidad citoplasmática, es posible que tenga su propia regulación, la cual puede ser independiente o estar en conjunción con la regulación de los filamentos de actina. Por ejemplo, parece ser que uno de los mecanismos que sirve para activar a la miosina ATPasa (el enzima encargado de hidrolizar el grupo fosfato terminal del ATP con el fin de proporcionar la energía necesaria para la contracción del sistema actina-miosina) es la modificación del propio enzima. David J. Hartshorne y colaboradores, de la Carnegie-Mellon University, y Adelstein y colaboradores descubrieron que la miosina ATPasa procedente de células muscula-



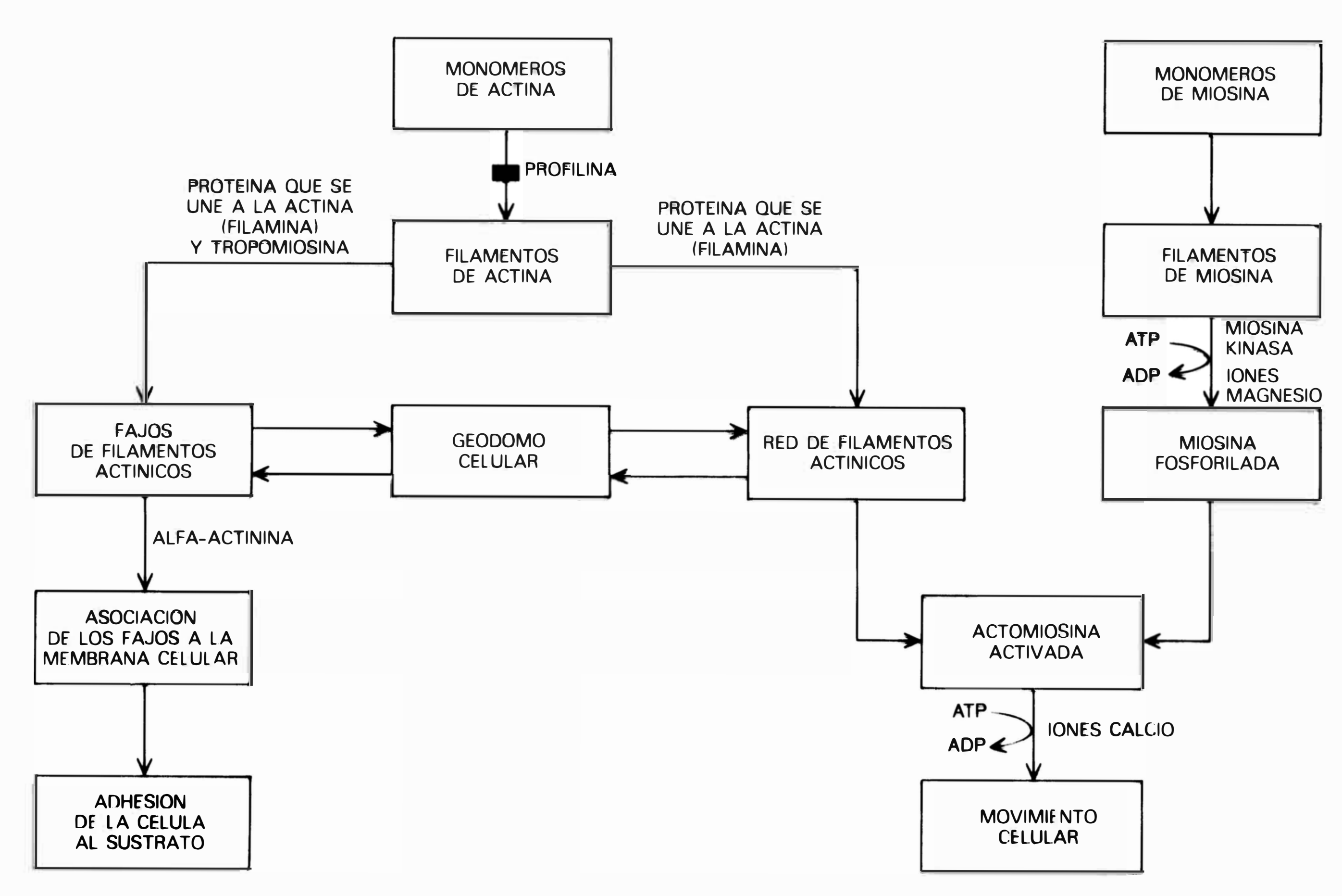
de actina. La mayoría de los hallados en una célula móvil, redondeada, están formando una red difusa (*a*). Cuando la célula deja de migrar y empieza a aplanarse, algunos filamentos se organizan formando un geodomo perinuclear (*b*). Luego, los fajos de filamentos actínicos se reúnen en los vértices de esta malla tridimensional (*c*). La red se disgrega dejando, en la célula ya totalmente aplanada, tan sólo fajos lineales de filamentos (*d*).

res y de plaquetas humanas tan sólo podía ser activada por la actina filamentosa una vez se hubiera llevado a cabo la fosforilación de la miosina por un enzima específico: una kinasa. Comparten este sistema regulador las células musculares, los fibroblastos y los mioblastos. Adelstein y sus colaboradores han intentado profundizar más en este campo y han descubierto en las células no musculares un segundo enzima: una fosfatasa que elimina de la miosina el fosfato que había sido unido a ella por la kinasa. La miosina desfosforilada no sigue conservando su capacidad para interaccionar con la actina y se inhibe la activación de la ATPasa por la actina. Por tanto, este sistema kinasa-fosfatasa permite una regulación de "quita y pon" de los fenómenos de motilidad que dependen de la miosina.

Por otra parte, el análisis bioquímico de la miosina muscular y la no muscular ha puesto recientemente en evidencia que un determiando tipo de célula puede poseer dos o más miosinas diferentes. Por consiguiente, si se tiene en cuenta el hecho de que la célula contiene un mínimo de dos y un máximo virtual de cinco diferentes tipos de actina, se comprende que estas moléculas pueden formar de cuatro a 10 combinaciones distintas. Además, algunas actinas se polimerizan en presencia de iones calcio o magnesio mientras otras sólo en ausencia de estos iones. Por todo ello se hace evidente que la célula posee un amplio repertorio de posibles interacciones actina-miosina entre las cuales puede escoger para sus diferentes actividades motoras.

Resumiendo, las células no musculares presentan mecanismos de control que regulan la actina monomérica, la actina en forma de fajos de filamentos, la actina en forma de gel o red y la actividad enzimática de la miosina. Se puede decir con certeza que algunas de las actividades motoras se originan por interacciones actina-miosina mientras otras quizá se desencadenen por una polimerización explosiva de la actina, una desorganización brusca de los filamentos de actina o una interconversión de los filamentos entre la forma de fajos y la de red. En un sistema dinámico, como es un caso típico el propio citoplasma, coexisten todos estos modos de regulación de la motiliad celular y del mantenimiento de la estructura citoplasmática, operando muchos de ellos de forma simultánea.

El conjunto de estos resultados demuestra hasta qué punto la regulación de la motilidad celular se desarrolla con múltiples fases y a diferentes niveles; aun y así, hasta ahora no se ha hecho más que rozar la superficie del problema. Es de esperar que el futuro aporte un gran número de enfoques nuevos, pues una aproximación ampliamente interdisciplinaria a la cuestión debería conducir a un conocimiento en profundidad de este fenómeno fundamental de la célula.



REGULACION DEL MOVIMIENTO CELULAR a nivel de moléculas en las células no musculares, esquematizada en este diagrama de flujo. La actina queda regulada a dos niveles: la polimerización de los monómeros de actina para formar filamentos está inhibida por la proteína reguladora profilina; por otra parte, hay otras dos proteínas reguladoras, la proteína que se une a la actina (o filamina) y la tropomiosina, encargadas de favorecer la ordenación de los filamentos de actina para formar un gel amorfo y fajos de elevada organización, respectivamente. Existe una cuarta proteína acompañante, la alfa-actinina, que participa en el proceso de asociación de los fajos de filamentos con la membrana celular. La miosina, en cambio, se regula principalmente por modificaciones químicas. Existe un enzima, la miosina kinasa, que transfiere el grupo fosfato terminal del ATP a la miosina, activando así a dicha proteína contráctil. Posteriormente, la miosina activada interacciona con los filamentos de actina del gel citoplasmático, dando origen al movimiento celular. Sin embargo, algunos tipos de movimiento celular no parecen requerir la presencia de miosina y quizá se originen, de un modo alternativo, por una rápida ordenación o desorganización de los filamentos de actina o mediante la intervención de microtúbulos.

Utilización del láser en química

La utilización de un láser permite suministrar energía con la frecuencia necesaria para conseguir que los productos reaccionantes lleven a cabo una determinada reacción química en la forma más conveniente posible

Avigdor M. Ronn

La química es la ciencia que se ocupa de las transformaciones moleculares. Una característica fundamental de estas transformaciones es la ganancia o la pérdida de determinadas cantidades de energía. Los cambios químicos no se pueden interpretar a un nivel fundamental y no pueden conducir a aplicaciones prácticas a menos que las cantidades específicas de energía que acompañan a las reacciones químicas sean conocidas y controladas. Cuando los átomos o las moléculas que contiene un recipiente se excitan convencionalmente utilizando calor o presiones elevadas, la reacción que tiene lugar transcurre según varias etapas que, además de la sustancia deseada, producen numerosos subproductos. Para conseguir que la reacción transcurra en una dirección concreta, el químico recurre con frecuencia a un catalizador. Estas sustancias pueden presentarse bajo formas muy diferentes. Por ejemplo, un catalizador puede ser un metal, un no metal o un compuesto organometálico; también puede ser un sólido, un líquido o un gas.

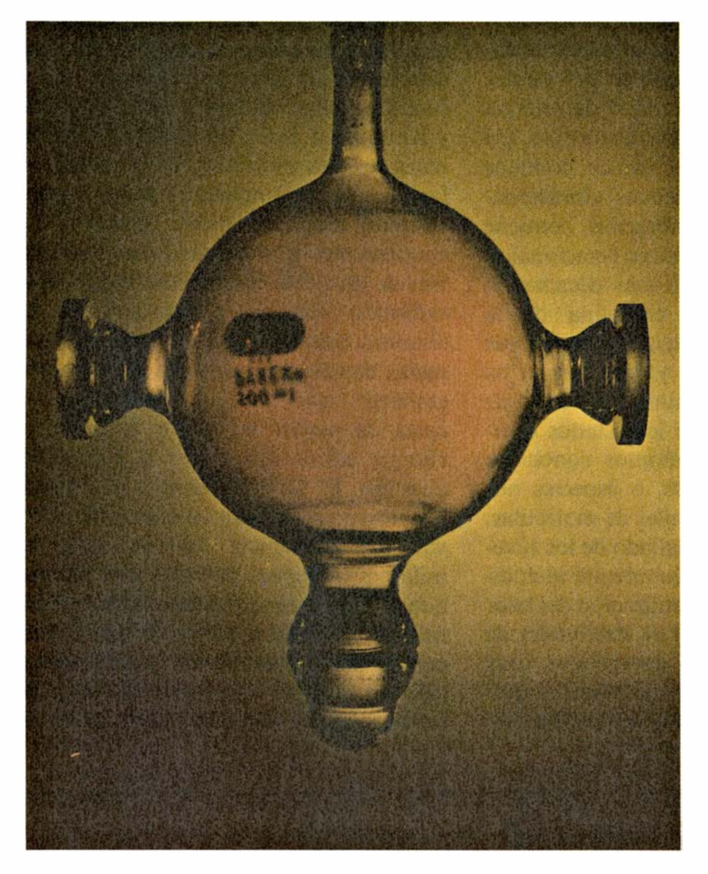
No está muy lejano el día en que la búsqueda empírica de una combinación óptima de temperatura, presión y materiales catalíticos para obtener un determinado resultado químico nos parecerá tan primitivo como la alquimia. Puesto que uno de los objetivos básicos de la química consiste en debilitar o romper unos determinados enlaces químicos, como preludio de una reacción concreta, cabe pensar en la posibilidad de que la energía necesaria sea aportada en el nivel o frecuencia que resulten más adecuados para conseguir este fin. Este control explícito de los cambios energéticos se está haciendo posible gracias a la utilización del láser. Los láseres proporcionan fuentes de energía extremadamente intensas para prácticamente cualquier longitud de onda de las regiones del ultravioleta, visible, infrarrojo o microondas. Las propiedades características del láser (monocromaticidad y alta intensidad) son ideales para el control de los cambios específicos de energía que provocan o catalizan las reacciones químicas.

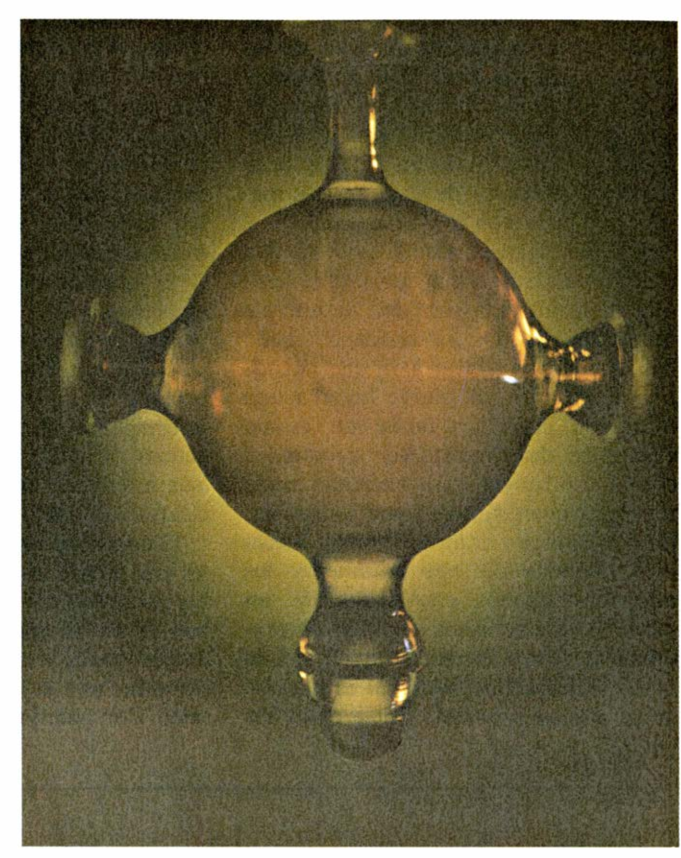
Con el fin de apreciar el impacto que los láseres han producido ya en la química y poder especular sobre el futuro que se abre ante nosotros en este sentido, es conveniente estudiar más detenidamente la forma en que la energía de las moléculas, los átomos y los electrones resulta afectada por la aportación de la energía de un láser. Vamos a ver cómo las limitaciones propias de la tecnología química convencional se pueden superar o evitar con facilidad mediante un sistema de láser adecuado.

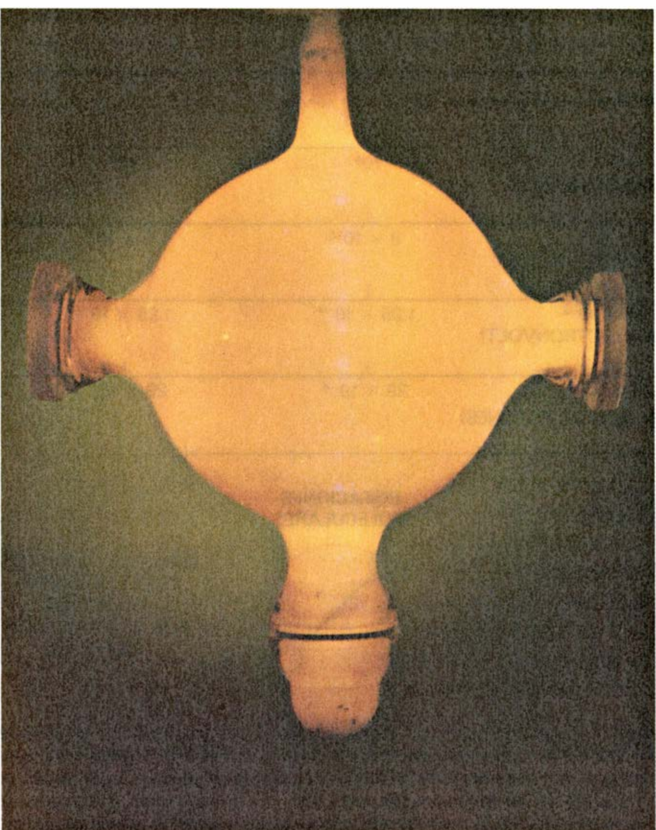
Las reacciones químicas pueden ocurrir y de hecho ocurren en una amplia variedad de medios y bajo condiciones muy diferentes. Evidentemente, ciertas reacciones químicas transcurren espontáneamente, tan pronto como los correspondientes gases, soluciones o sólidos se ponen en contacto. Estas reacciones no requieren un aporte de energía desde el exterior y, por tanto, no resultan adecuadas para la utilización del láser. Otras reacciones, y probablemente la mayoría de las reacciones químicas de interés industrial en la actualidad y en el futuro, deben contar con la acción de calor, presión o catalizadores, o una combinación de todos estos agentes. La eficiencia con que dichos factores se acoplan a los átomos o moléculas reaccionantes determina el rendimiento de los productos de las correspondientes reacciones y, en consecuencia, la eficiencia de tales reacciones. Nuestro objetivo radica en suministrar energía directamente al canal de reacción adecuado, en lugar de hacerlo al azar, desperdiciándola en los muchos canales de reacción posibles que existen en cada caso. Puesto que el láser es capaz de suministrar cantidades importantes de energía con una determinada frecuencia o nivel energético, cuyo valor puede corresponder a un canal de reacción en particular, se está convirtiendo en una nueva herramienta de precisión para producir transformaciones químicas.

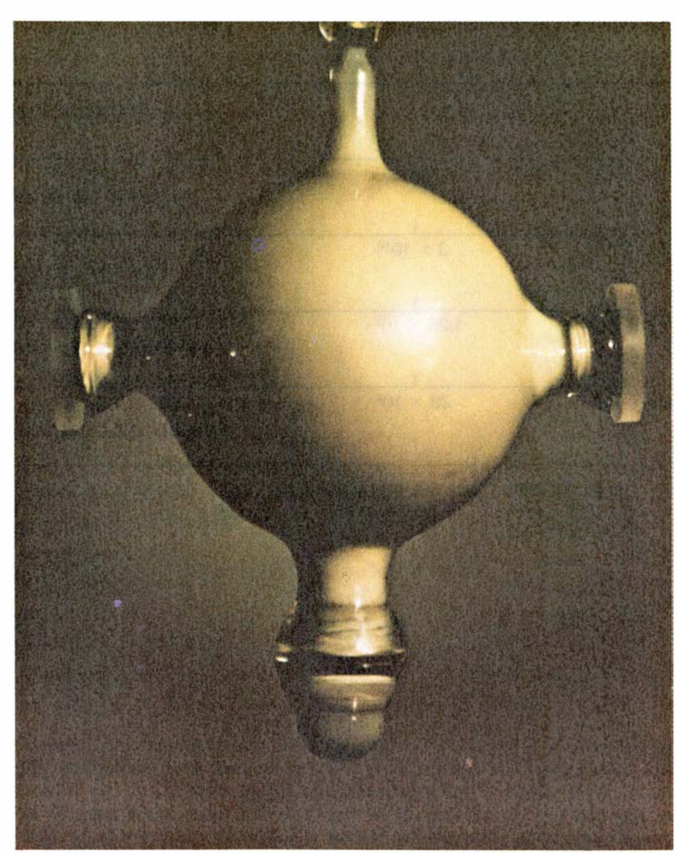
Cada una de las regiones principales del espectro electromagnético a las que se puede ajustar el láser, desde el ultravioleta al infrarrojo, presenta unas determinadas propiedades en relación con la forma en que la energía se puede suministrar a un sistema reaccionante. Una de las propiedades fundamentales de los átomos y de las moléculas consiste en que estos cuerpos emiten o absorben energía sólo en cantidades cuantificadas o discretas. Cuando un átomo o una molécula absorbe un cuanto de energía sufre una transición desde un estado de menor energía hasta otro estado de energía mayor. El salto puede ser pequeño o grande y afectará a diferentes propiedades del átomo o molécula según el valor de la energía del cuanto.

La absorción de un cuanto de radiación visible o ultravioleta altera el estado de excitación de la estructura electrónica de un átomo o de una molécula. La estructura electrónica describe la configuración y el tamaño de las órbitas de los electrones que se mueven alrededor del núcleo central de un átomo o, en el caso de una molécula, alrededor de los núcleos de dos o más átomos. La absorción de un cuanto de energía correspondiente a radiación visible o ultravioleta se traduce en la elevación de un electrón a un nivel superior. Por otra parte, la absorción de un cuanto de radiación infrarroja altera los niveles energéticos del sistema representado por los átomos de la molécula. Estos cambios consisten en vibraciones, es decir, en un movimiento relativo de uno o más átomos respecto al centro de masas de la molécula. Cuando un átomo individual absorbe un cuanto de radiación infrarroja, simplemente se mueve más deprisa; generalmente, nos referimos a esta situación diciendo que la energía de traslación del átomo ha aumentado. La energía que transporta un cuanto de la región espectral de las microondas es del orden de 10^4 veces menor que la que corresponde a un cuanto de la radiación







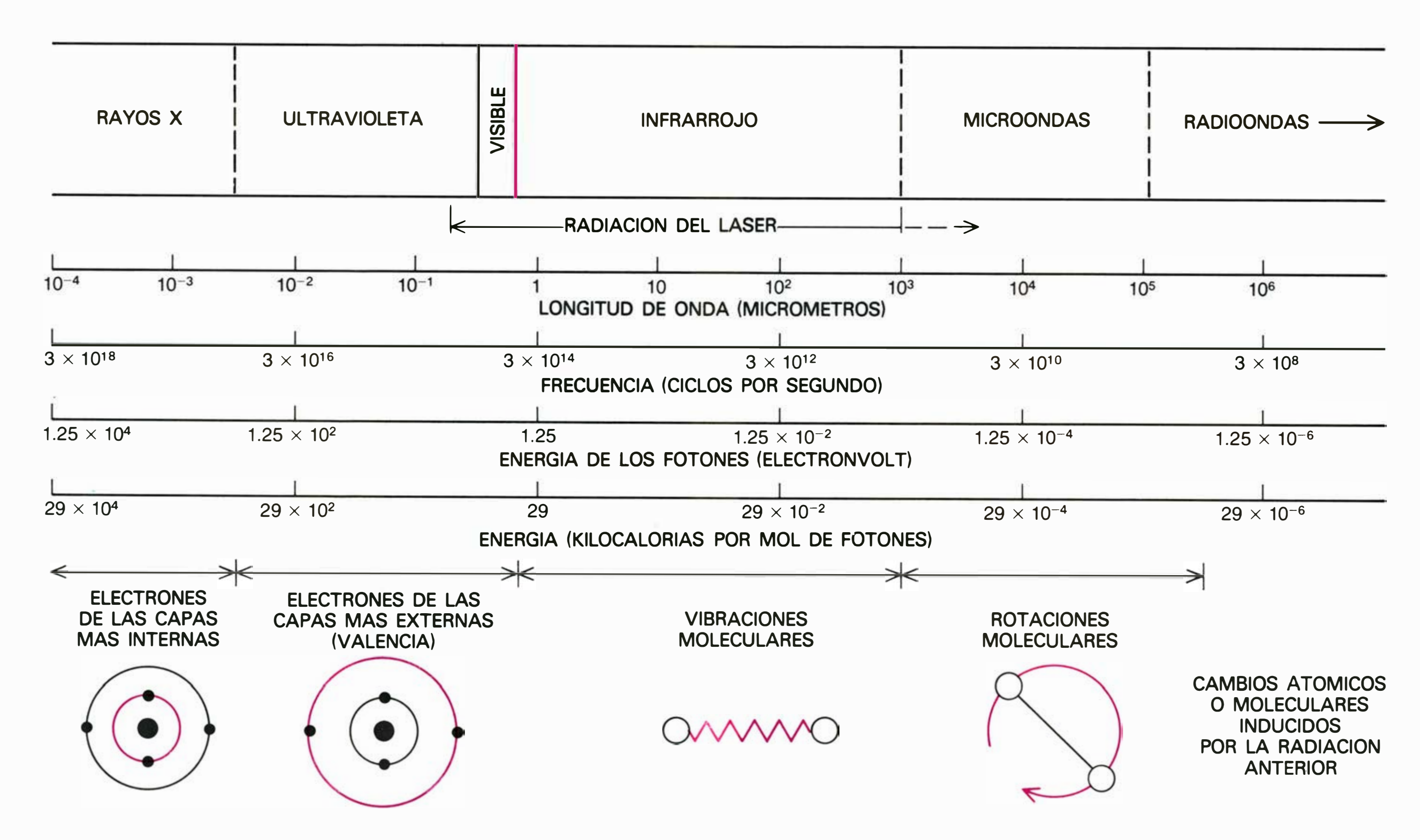


DESCOMPOSICION DEL CLORURO DE CROMILO inducida por un haz de fotones infrarrojos procedentes de un láser de dióxido de carbono. El cloruro de cromilo (CrO_2Cl_2), un gas de color castaño rojizo a temperatura ambiente, se aprecia en el recipiente que se reproduce en la fotografía superior izquierda antes de ser sometido al haz láser. El gas está sometido a una presión de 0,005 atmósferas. La fotografía de la parte superior derecha se tomó en el momento en que el haz de láser, consistente en fotones cuya longitud de onda es de 10 micrometros, penetra en el citado recipiente. El haz rojo está formado por las moléculas del cloruro de cromilo que han absorbido de 16 a 18 fotones en rápida sucesión y han sido llevadas a un estado excitado electrónicamente. Algunas de las moléculas excitadas liberan un fotón de luz visible, incorporando toda la energía de los 16 a 18 fotones infrarrojos absorbidos, y regresan al estado fundamental original. Cuando el hidrógeno está presente junto al cloruro de cromilo se produce una reacción diferente. Al absorber unos 28 fotones, el CrO_2Cl_2 se descompone en CrO_2Cl y un átomo de cloro. Las reacciones en cadena que tienen lugar en forma subsiguiente a esta descomposición provocan una explosión (*inferior izquierda*). El recipiente de la fotografía situada en la parte inferior derecha está lleno de partículas de óxido de cromo (Cr_2O_3), cloruro de hidrógeno (ClH) y vapor de agua. (Fotografías tomadas por Fritz Goro.)

visible, pero aun así la absorción de un cuanto de microondas es capaz de aumentar la energía de traslación de un átomo o una molécula. Si el absorbente es una molécula, el cuanto de microondas puede hacer también que el movimiento de rotación de los átomos de la molécula alrededor de su centro de masas transcurra con mayor velocidad.

En consecuencia, resulta evidente que una fuente de láser puede suministrar a una molécula una cantidad de energía perfectamente cuantificada (dependiendo de la región espectral a la que corresponda su emisión), dando lugar a una transición específica desde un estado energético determinado (en general, el estado fundamental no excitado) hasta un estado energético de energía superior. De acuerdo con la interpretación mecánico cuántica de este fenómeno, los diferentes estados energéticos de un átomo o una molécula se pueden representar en forma de una serie vertical de líneas horizontales [*véase la ilustración de la página opuesta*]. Entre todos los niveles electrónicos, o niveles asociados a los estados de excitación de los electrones, es en donde se da la mayor separación entre niveles, que se corresponden con las mayor cantidad de energía por cada cuanto emitido o absorbido. Un determinado nivel electrónico contiene muchos estados energéticos vibracionales; cada nivel de vibración contiene muchos estados energéticos rotacionales. A lo largo de estas últimas décadas, la espectroscopía —una disciplina de la física y la química dedicada al estudio de la estructura atómica y molecular— ha proporcionado una enorme cantidad de información relativa a los estados energéticos de todos los átomos conocidos (sus diferentes isótopos, o especies nucleares), así como de miles de moléculas.

El conocimiento detallado de los niveles energéticos es absolutamente ineludible para proceder a la utilización del láser con el fin de producir la excitación de determinados estados energéticos para provocar una determinada reacción química. Con anterioridad a la introducción del láser, la única opción de que disponía el químico para proporcionar energía a una reacción química, consistía en usar una fuente de calor convencional; estas fuentes excitan de forma indiscriminada y simultánea todos los estados vibracionales de una molécula. Evidentemente, las fuentes convencionales de luz y calor —llamas, arcos eléctricos, lámparas incandescentes y similares— son susceptibles de un cierto control haciendo que la radiación que emiten pase a través de sistemas capaces de seleccionar determinadas bandas de longitudes de onda; sin embargo, este resultado se consigue a costa de perder la mayor parte de la energía suministrada por la fuente en cuestión. El flujo de energía que puede proporcionar un láser en cualquier intervalo de frecuencia que se desee es millones o billones de veces más intenso que el flujo que se puede obtener en ese mismo intervalo a partir de una fuente convencional recurriendo a una operación de filtrado. Así, hasta un láser de laboratorio de tamaño modesto emite



RESPUESTA ATOMICA Y MOLECULAR a la radiación electromagnética. Dicha respuesta depende de la longitud de onda de la radiación y, por tanto, de la cantidad de energía que transporta cada fotón. Los fotones extremadamente energéticos de la región de los rayos X del espectro de ondas electromagnéticas pueden alterar el estado de excitación de los electrones más internos de un átomo. Los fotones de energía menor correspondientes a las regiones ultravioletas y visible del espectro interaccionan con los electrones de las capas externas del átomo, es decir, precisamente aquellos electrones que toman parte en las reacciones químicas. Cuando un átomo absorbe un fotón o cuanto de la radiación ultravioleta o visible, el correspondiente electrón sufre una transición a una órbita de mayor energía que la inicial. En general, los fotones infrarrojos producen una alteración en los estados de vibración de los átomos que constituyen una molécula. Los fotones de la región de las microondas, cuya energía es todavía menor, afectan a los estados rotacionales de las moléculas. Dado que los láseres se pueden ajustar de forma tal que la frecuencia (y, por tanto, la energía) de los fotones que proporcionan tenga un valor exactamente determinado, pueden provocar la producción de ciertos cambios muy precisos en el estado de excitación de los átomos y las moléculas. En consecuencia, los láseres se han convertido en la herramienta adecuada para crear el estado de excitación que resulte más favorable para dar lugar a una reacción química en particular. Actualmente los láseres son capaces de proporcionar prácticamente cualquier longitud de onda comprendida entre los 0,2 micrometros (en el ultravioleta) y los 100 micrometros (en el infrarrojo). Los láseres pueden también producir microondas, aunque en este caso se suelen denominar máseres. A los químicos les puede interesar expresar el contenido de energía de la radiación producida por un láser en términos de kilocalorías por mol. Un mol es el peso en gramos de una sustancia igual al peso molecular (o al peso atómico) de dicha sustancia y consiste en 6,02 × 10^{23} moléculas (o átomos) de la misma. Mol de fotones = 6,02 × 10^{23} fotones.

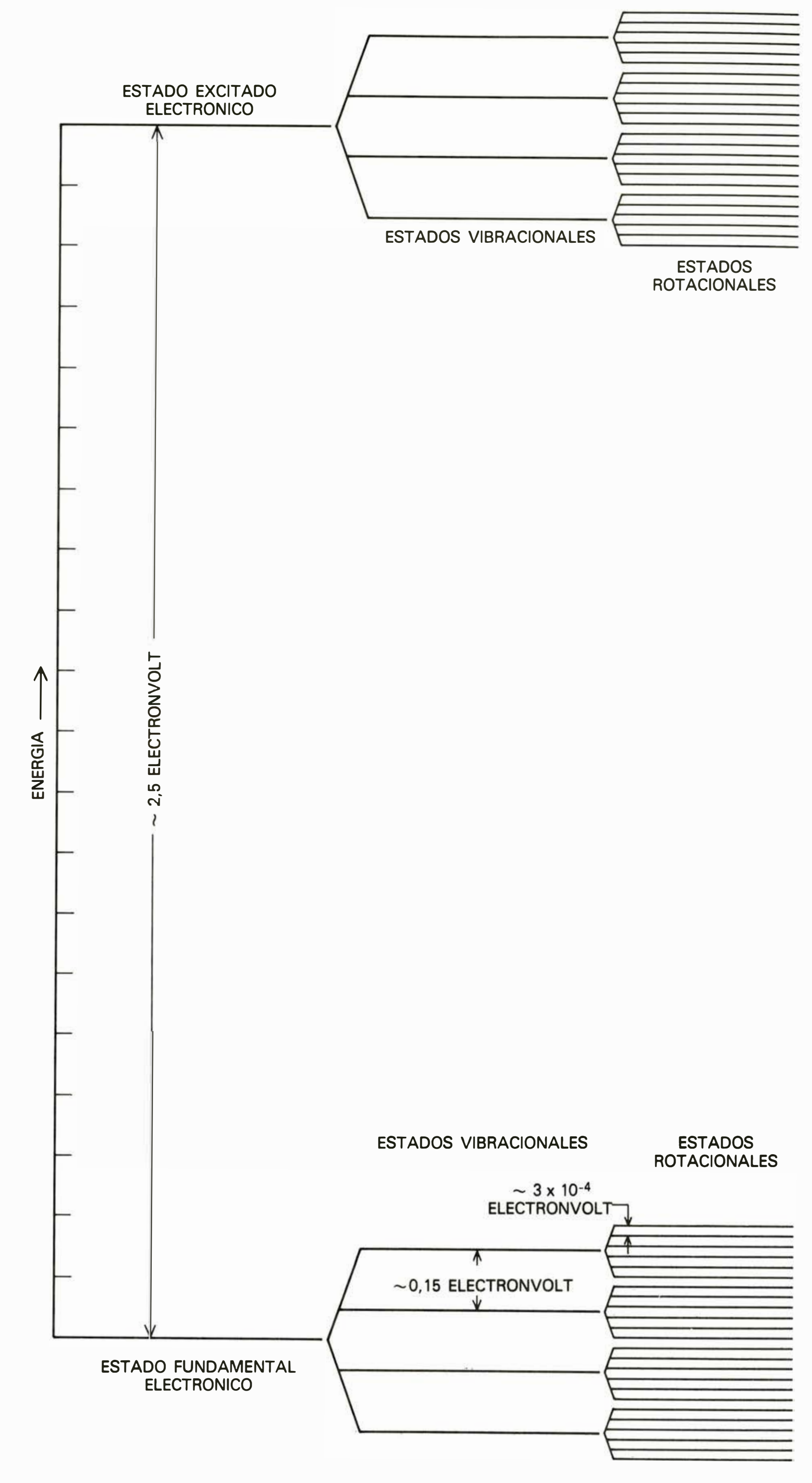
muchos más fotones por unidad de superficie de su haz que los que emite el sol también por unidad de superficie para la misma longitud de onda.

En virtud de lo que acabamos de decir, podría sacarse la impresión de que la introducción del láser es capaz de satisfacer todos los sueños de los químicos, proporcionándoles el medio de excitar selectivamente a los reaccionantes –electrónica, vibracional o rotacionalmente, de acuerdo con sus deseos– y controlar una reacción en la forma en que elija previamente, sin que tenga lugar la formación de subproductos. Sin embargo, en la práctica, la química del láser resulta afectada frecuentemente por ciertas restricciones que deben ser superadas antes de que ese sueño se convierta en realidad.

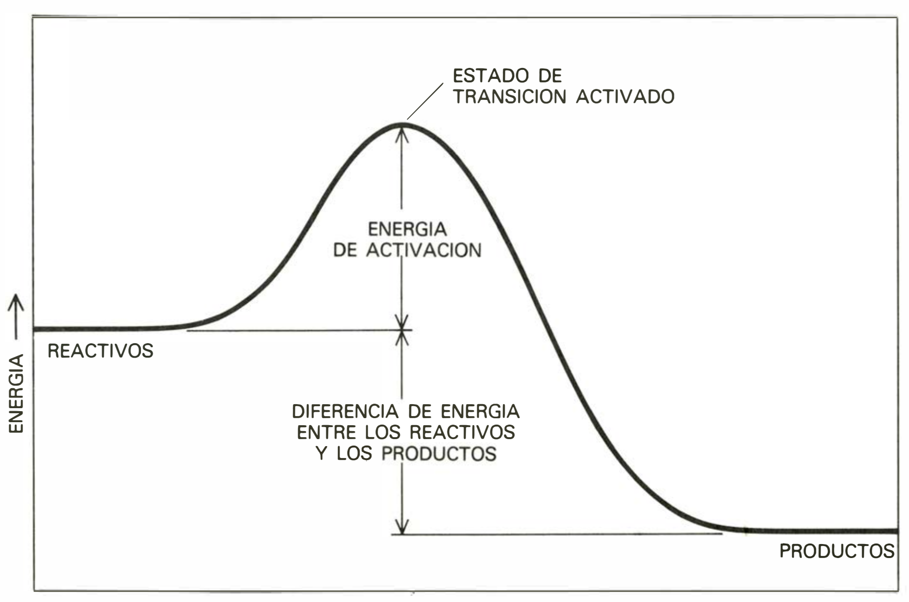
Un átomo o una molécula puede absorber radiación sólo en forma de cuantos discretos, y puede hacerlo así sólo para aquellos estados energéticos que permite la mecánica cuántica. Las absorciones secuenciales de fotones pueden llevar a un átomo o a una molécula hasta niveles energéticos extremadamente elevados. En última instancia, la energía absorbida provoca la extracción de uno o más electrones del átomo en cuestión; en este caso, se dice que el átomo está ionizado. De forma análoga, una molécula se disociará en fragmentos o incluso en sus átomos constituyentes. Desde el punto de vista de la mecánica cuántica, se puede considerar que la ionización y la disociación conducen a un estado en el que existe un continuo de niveles energéticos permitidos en lugar de niveles discretos.

La cuestión fundamental que se le plantea al químico que utiliza un láser es saber qué sucede con el exceso de energía que ha sido depositado selectivamente por el láser en un átomo o molécula durante el intervalo de tiempo que debe transcurrir hasta que la reacción tenga lugar. Si la energía se disipa con excesiva rapidez a través de canales no productivos, la reacción no se producirá nunca o, si lo hace, progresará muy lentamente.

Un átomo o una molécula puede perder su excitación a través de una serie de mecanismos diferentes. Ante todo, puede tener lugar simplemente la reemisión de un fotón con la misma energía que la del fotón absorbido originalmente, con lo cual el átomo o la molécula en cuestión retornan a su estado inicial. Como alternativa, dichos cuerpos pueden liberarse de su energía en exceso a través de una o más colisiones con los átomos o las moléculas vecinos. En este caso, el exceso en



ESTA IMAGEN MECANICO-CUANTICA se refiere al hecho de que una molécula puede ocupar solamente ciertos estados energéticos discretos. En la configuración de estos estados intervienen las órbitas de los electrones, las vibraciones de los átomos individuales que constituyen una molécula y las rotaciones de las moléculas como un todo. Para cada estado electrónico, existen muchos estados vibracionales posibles y, para cada estado vibracional, existen muchos estados rotacionales posibles. Por ejemplo, la absorción de un fotón de microondas de baja energía en el estado fundamental es suficiente para elevar a una molécula hasta un estado rotacional de energía mayor. Para elevar a una molécula desde el estado fundamental a un estado electrónico excitado se necesita un fotón de radiación visible o ultravioleta. Cuando una molécula sufre una transición desde un estado a otro de energía menor, se produce la emisión de un fotón cuya energía tiene un valor perfectamente determinado.



PARA QUE SE PRODUZCA UNA REACCION QUIMICA se necesita generalmente la aportación de una cierta cantidad de energía. Esta figura representa la conversión de un reactivo simple en uno o más productos de reacción. La energía de activación se aporta con el objeto de elevar el reactivo hasta un estado de transición elevado. Una vez que dicho reactivo ha alcanzado el citado estado, es muy probable que descienda de la colina superando la barrera de energía y, en consecuencia, sufra una descomposición. Sin embargo, también existe cierta probabilidad de que se libere de su recién adquirida energía y regrese a su estado original. La figura se refiere a una reacción exotérmica, es decir, a un proceso en cuya realización se produce una liberación de energía. En las reacciones bimoleculares, en las que dos sustancias dan lugar a la formación de nuevos productos de reacción, se sigue un camino energético similar al descrito. Una reacción está catalizada por láser si este aporta la "chispa" inicial.

cuestión o bien se transfiere de golpe en un solo cuanto o bien se emite en forma secuencial mediante un cierto número de cuantos de energía menor. Como ya hemos dicho anteriormente, el sistema puede ionizarse o disociarse si la energía recibida es lo suficientemente grande para ello. Si la ionización o la disociación tiene lugar antes de que los átomos o las moléculas sean capaces de colisionar, se pueden presentar algunas perspectivas muy interesantes.

Vamos a considerar lo que puede sucederle a una molécula al absorber un cuanto de infrarrojo capaz de llevarla desde un estado de vibración de baja energía a otro de alta energía. Si el químico expone una especie molecular en fase gaseosa a baja presión al haz de un láser de infrarrojos, puede medir con bastante precisión el curso temporal del flujo de energía entrante y saliente de los estados energéticos susceptibles de ser excitados. Estas experiencias se conocen con el nombre de estudios de relajación vibracional. En la mayoría de los sistemas analizados hasta la actualidad (varios centenares), una cierta fracción de las moléculas irradiadas son elevadas rápidamente hasta un nivel energético particular, a partir del cual la energía se utiliza en descomponer a la molécula o se disipa dentro de esa misma molécula a través de una serie de colisiones. Por otra parte, las colisiones pueden disipar la energía inicial entre las moléculas próximas. Como consecuencia de todos estos procesos resultan excitados muchos estados energéticos de vibración.

En particular, la redistribución de la energía favorece a dos canales de reacción concretos. En uno de ellos, la energía se transfiere desde el estado excitado de vibración hasta uno o más de los niveles de vibración de la molécula mediante colisiones. El proceso continúa hasta que toda la energía de vibración se ha transferido a niveles rotacionales y de traslación del estado fundamental de vibración. En el segundo tipo de redistribución, el exceso de energía del nivel excitado inicialmente se transfiere directamente a los estados rotacionales y traslacionales. En general, el primer tipo de redistribución de la energía (la transferencia de energía de vibración a otros estados vibracionales dentro de la molécula) tiene lugar entre diez y varios miles de veces más de prisa que la degradación en una sola etapa desde un estado de vibración hasta un estado rotacional o traslacional.

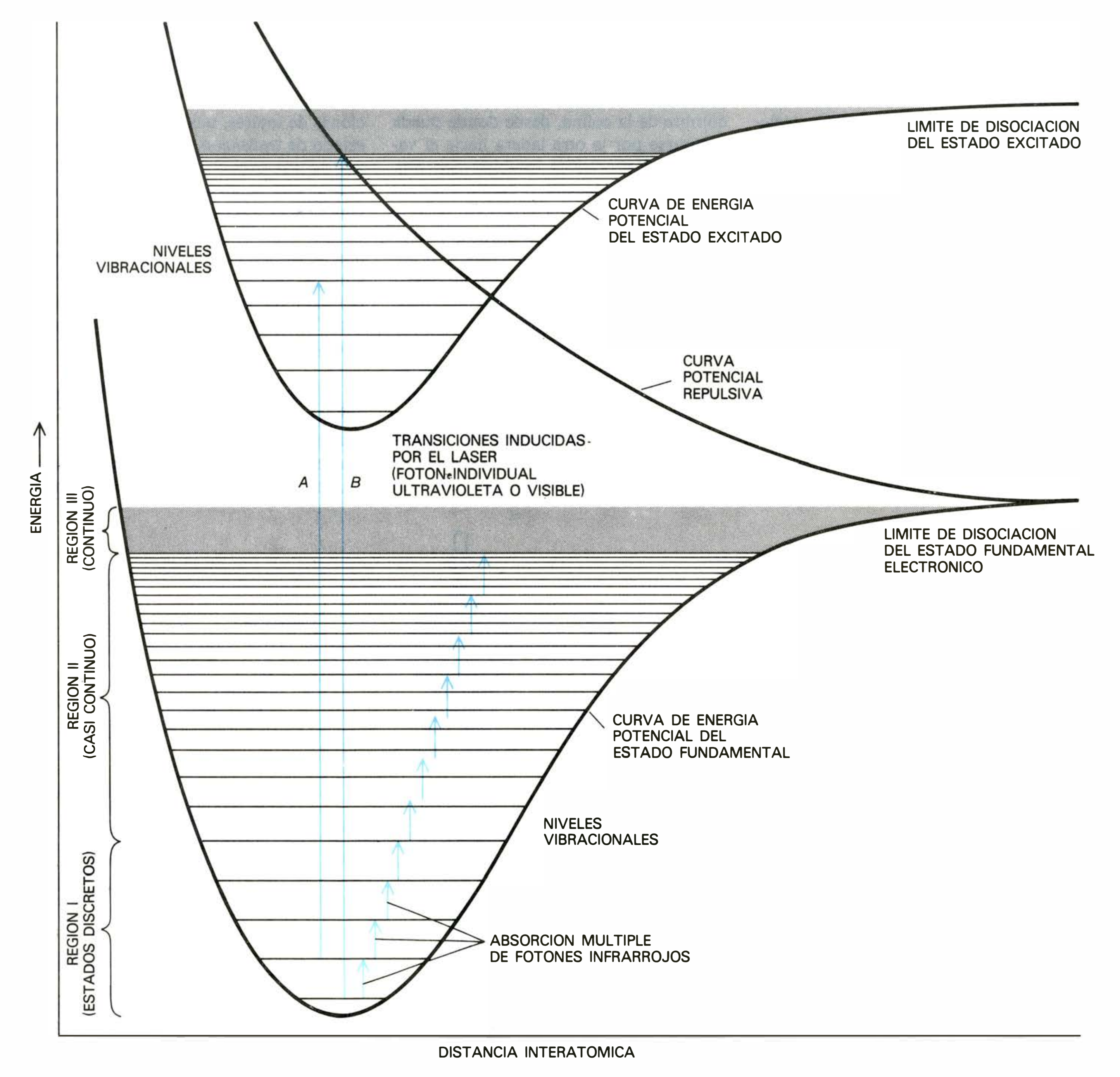
En consecuencia, se pueden concebir tres formas diferentes de llevar a cabo una experiencia con láser. En la primera de ellas, la reacción en cuestión transcurre en un tiempo menor o igual que el tiempo necesario para que se produzca la transferencia de energía de vibración desde el estado vibracional excitado inicialmente hasta cualquier otro estado de vibración de la misma molécula. En el segundo tipo de experiencia, la reacción transcurre a lo largo de un tiempo considerablemente mayor, compatible con el tiempo que dura la transferencia de energía desde el estado de vibración inicial hasta los estados rotacionales y traslacionales de energía menor. El tercer tipo de experiencia reviste una importancia enorme, puesto que se refiere a la disociación molecular (o ionización atómica) que tiene lugar antes de que se produzca cualquier degradación de energía en virtud de colisiones moleculares (o atómicas). Bajo tales condiciones, la energía del láser se utiliza en su totalidad y de forma selectiva para provocar la producción de unas determinadas reacciones.

Evidentemente, las reacciones del primer tipo, las que transcurren tras la excitación de un único estado vibracional, resultan particularmente adecuadas a la activación con láser debido a que resulta excitado de forma preferente un camino de reacción único. En el segundo tipo de reacción, en el que resultan excitados todos los estados de vibración, no cabe esperar un alto grado de selectividad. En consecuencia, el experimentador debe escoger la secuencia de su reacción una vez haya establecido el marco temporal para la transferencia de energía de vibración desde el estado inicial hasta los restantes estados de vibración y, finalmente, hasta los estados de rotación y de traslación propios de un determinado sistema. Se puede considerar que el segundo tipo de experiencia al que antes hemos aludido es básicamente de "calentamiento", debido a que la reacción depende no tanto de un estado de vibración específico, sino más bien de todos los estados de vibración de alta energía. Esta condición se puede alcanzar igualmente bien mediante un calentamiento convencional. La diferencia principal radica en el tiempo que se requiere para alcanzar una excitación total. Con la ayuda de un láser, dicho tiempo puede ser inferior al microsegundo, en comparación con el segundo o más que se necesita si se recurre a una fuente térmica convencional. Un tiempo pequeño puede ser un factor importante para muchas reacciones de interés debido a que esta circunstancia puede motivar la existencia o no de diferentes estados intermedios en el transcurso de las mismas.

Vamos a considerar ahora los dos tipos de reacciones cuyo estudio resulta ventajoso en fase gaseosa: las reacciones en las que intervienen únicamente moléculas de un solo tipo y las reacciones en las que intervienen moléculas de dos

tipos distintos. En la reacción unimolecular, después de haber sido excitado recurriendo a un procedimiento convencional o mediante un láser, el reactivo se transforma en un nuevo producto o se descompone en dos o más productos, con o sin colisiones moleculares. En una reacción bimolecular, existe la posibilidad de producir la excitación de uno o de los dos reactivos de forma que, a continuación, las colisiones entre ellos den lugar a la formación de un nuevo producto.

Los dos tipos de reacción nos llevan a las dos formas diferentes de actuación del láser sobre las reacciones químicas: la inducción y la catálisis. Cuando hablamos de la inducción de una reacción mediante el láser, no sólo nos referimos al hecho de que éste provoca su realización, sino que además suministra toda la energía termodinámicamente necesaria para conducir la reacción desde el principio al fin. Se dice que una reacción está catalizada por el láser si éste proporciona sólo la "chispa" inicial, tras de lo cual la reacción transcurre por sí misma sin necesitar una aportación posterior de energía. Evidentemente, las reacciones catalizadas, puesto que no hacen uso del láser como fuente de energía, son más fáciles de estudiar, comprender y evaluar. Por otra parte, las reacciones indu-



FRAGMENTACION MOLECULAR, que puede inducirse mediante un láser de varias formas diferentes. Las curvas de energía potencial representan la variación de la energía de enlace molecular en función de la distancia interatómica en el caso de una molécula típica. Cada curva posee muchos niveles vibracionales. Cuando una molécula absorbe un fotón de radiación visible (*A*) o ultravioleta (*B*) sufre una transición desde el estado fundamental electrónico hasta un estado excitado de mayor energía. La probabilidad de disociación aumenta a medida que los estados de vibración a los que se lleva a la molécula se acercan al límite de disociación. La curva del potencial repulsivo representa una configuración molecular inestable. Si tras sufrir una excitación, la molécula adquiere esta configuración, la disociación es inmediata. La disociación se puede producir también directamente a partir del estado fundamental electrónico por absorción múltiple de fotones infrarrojos. En la región discontinua de la curva (I), los fotones infrarrojos se pueden absorber únicamente si su energía se ajusta exactamente al intervalo cuántico existente entre los estados vibracionales. En la región cuasicontinua (II) la exigencia del ajuste de la energía es menos crítica. No existe ningún mecanismo mediante el cual la absorción de fotones infrarrojos pueda dar lugar a una disociación partiendo de un estado electrónico excitado, a no ser que el estado superior se solape con el estado inferior.

cidas por el láser son, con mucho, únicas y no pueden duplicarse por medios convencionales.

En general, las reacciones unimoleculares exigen un elevado aporte inicial de energía, antes de que el reactivo se convierta en un producto determinado. En consecuencia, las reacciones unimoleculares se inscriben más bien en la categoría de las reacciones que se pueden inducir por láser y no de las reacciones catalizadas por láser. Las reacciones bimoleculares se sitúan indistintamente en ambas categorías. El caso en el que sólo pretendemos que el láser suministre una pequeña cantidad de energía para acelerar una reacción que de forma natural tendría lugar mucho más lentamente es comparable a una catálisis química convencional: los fotones del láser actúan como catalizadores.

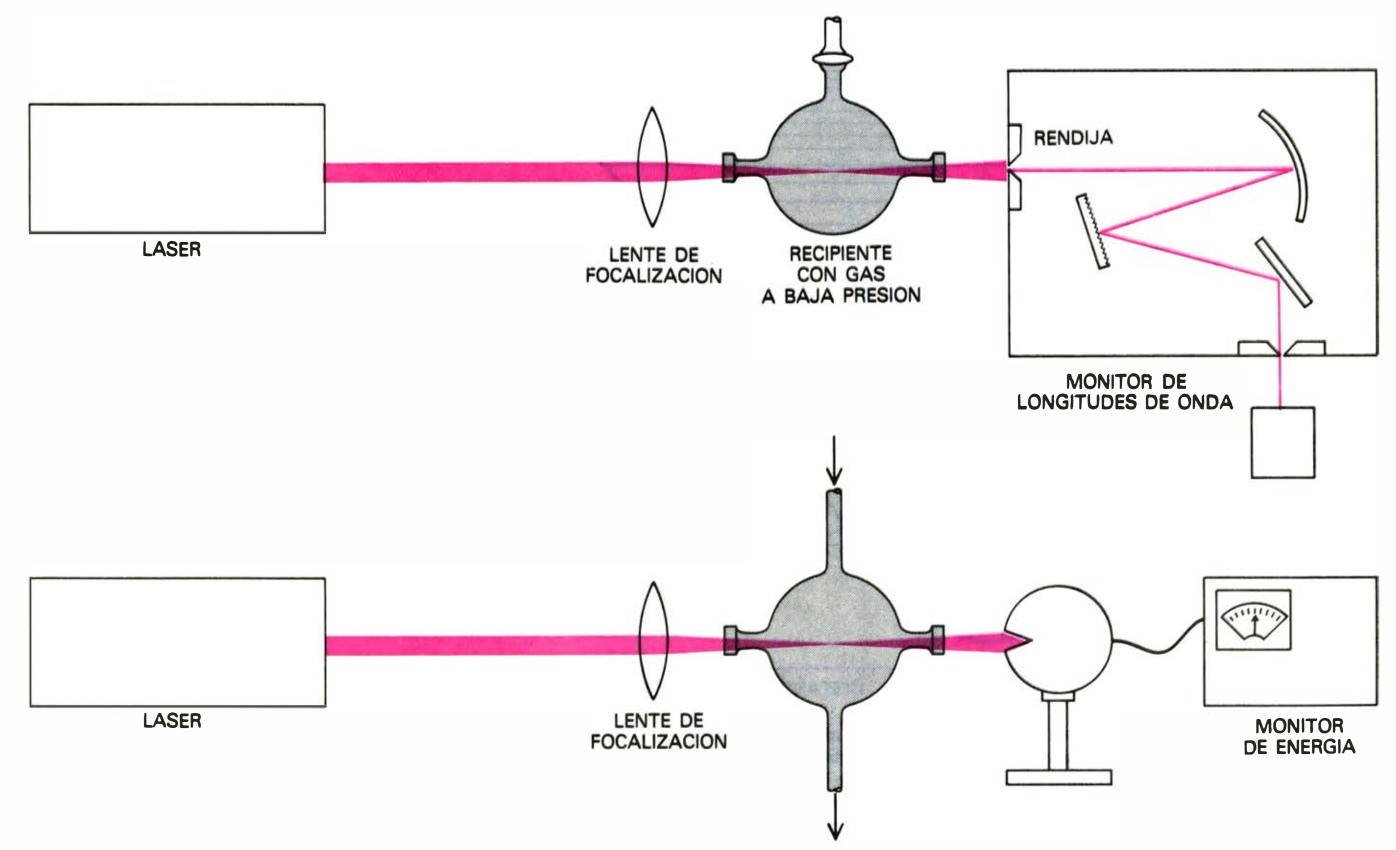
Desde el punto de vista energético, el camino que sigue una reacción química típica, según dicha reacción sea unimolecular o bimolecular, puede ilustrarse de una forma muy sencilla. Imaginemos una carretera que discurre sobre una colina vista de perfil. El principio de la carretera (que representa a los reactivos) está situado a la izquierda y le corresponde una cota, o nivel de energía, más alta que el final de la carretera, situada al pie de la colina y a la derecha (representando a los productos de la reacción). Como sucede siempre en química, los estados de energía más bajos son los más estables; en consecuencia, los productos estables poseen un contenido energético menor que el de los reactivos. La colina representa la energía de activación, entendiendo por tal la cantidad de energía que debe suministrarse al sistema para conducir la reacción a lo largo de un camino en particular.

Si la reacción es unimolecular, el reactivo en cuestión debe recibir la energía suficiente como para llevarlo a la cumbre de la colina, desde donde pueda deslizarse por la otra ladera hacia el valle de los productos estables. Sin embargo, debería tenerse presente que una vez que el reactivo ha alcanzado la cima de la colina, existe también cierta probabilidad de que regrese a su estado original. Si mediante algún proceso de desexcitación, pierde la energía que acaba de adquirir, el reactivo retornará a su estado inicial no excitado. Por el contrario, si prosigue su camino a lo largo de la pendiente, se transformará en sus productos.

Si la reacción es bimolecular, ambos reactivos deben alcanzar la cumbre de la colina. Cada reactivo puede elevarse hasta dicha cima absorbiendo energía independientemente, o bien uno de ellos puede tomar una cantidad de energía que sea suficiente para ambos y transferir luego parte de dicha energía a su compañero. En la cumbre de la barrera de energía, los dos reactivos forman un estado de transición a partir del cual existe una probabilidad finita de que retornen a sus estados originales y una probabilidad mucho mayor de que reaccionen entre sí y formen otros productos diferentes y más estables.

El estado de transición representa una configuración crítica. Para muchas reacciones de interés, nos gustaría saber si el estado de transición al que se llega a través de una excitación con láser es idéntico al que se alcanza por medios convencionales. La respuesta a esta cuestión no está todavía clara; la información de que se dispone actualmente en relación con las reacciones químicas inducidas por láser no es lo suficientemente abundante como para justificar una conclusión general.

A lo largo de estos últimos años se ha dedicado gran atención a las reacciones unimoleculares inducidas por láseres de infrarrojos. El aspecto más interesante de este problema radica en el



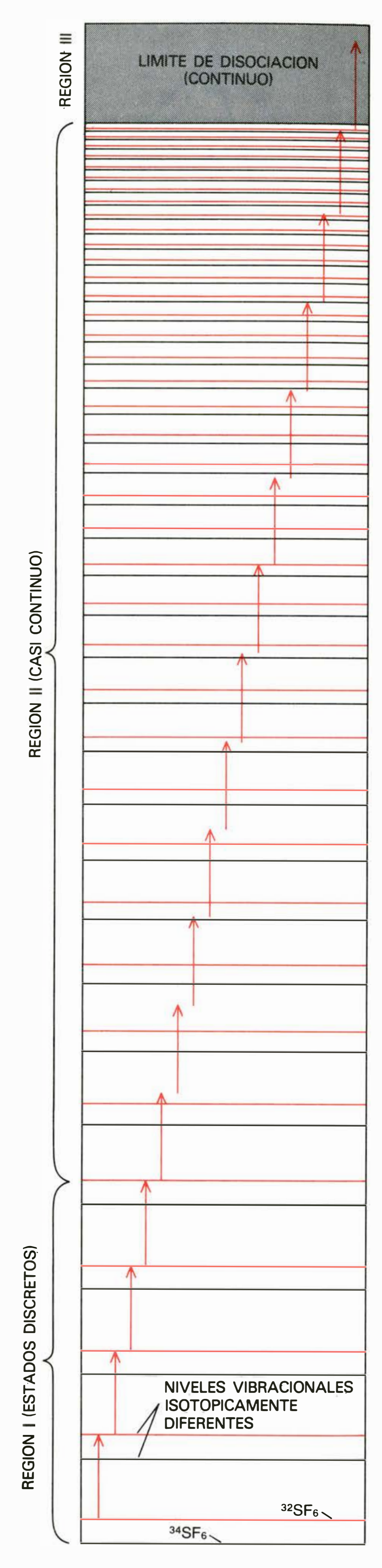
APARATOS DE LABORATORIO utilizados en el estudio de la química con la ayuda del láser. En uno de ellos (*arriba*) el material de partida se mantiene encerrado en un recipiente adecuado durante la irradiación con el haz del láser. En el segundo dispositivo (*abajo*) los reactivos se irradian a medida que fluyen a través de la cámara de reacción en forma continua. Con este segundo sistema la eficiencia de la reacción se puede ir estudiando mediante un análisis de los gases de salida. Se representan también dos de los dispositivos utilizados para controlar la radiación emergente del láser. En el dispositivo representado arriba, la correspondiente longitud de onda se determina con ayuda de una red de difracción. En el dispositivo reproducido en la parte inferior, se procede simplemente a recoger el haz y medir su energía sin determinar la longitud de onda de la respectiva radiación.

hecho de que la energía que transporta un fotón infrarrojo individual es insuficiente para disociar una molécula que se encuentre en su estado fundamental, o estado de más baja energía. Se puede considerar que la curva de potencial del estado fundamental está ocupada por muchos niveles vibracionales [*véase la ilustración de la página 73*]. Esta curva representa simplemente el hecho de que la molécula en cuestión está ligada y existe en el estado electrónico fundamental.

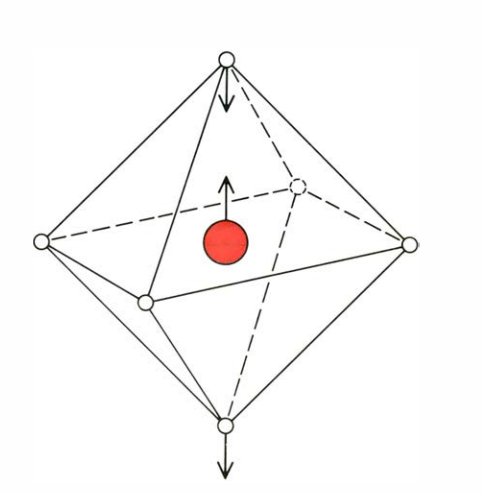
La absorción de un simple fotón infrarrojo sólo permite que la molécula suba un peldaño a lo largo de la escalera de la energía. En consecuencia, para que la molécula se disocie se deben absorber varios fotones de infrarrojos en serie secuencial. Por otra parte, si la molécula absorbe un fotón procedente de un láser de luz visible o ultravioleta, fotón mucho más energético que el anterior, puede ser llevada en una sola etapa desde el estado fundamental hasta un estado electrónico excitado situado bastante por encima del estado fundamental, a partir del cual se puede producir más fácilmente la disociación. En el caso de que exista la posibilidad de atravesar la que se conoce como curva de potencial repulsivo, la energía necesaria para que se produzca la disociación a partir de este estado resulta ser todavía menor. Dicha curva se refiere a la situación de la molécula en una configuración repulsiva y no en un estado ligado y, en consecuencia, la molécula se disocia aunque no disponga de toda la energía que se necesita normalmente para ello. La mayor parte del trabajo que se ha llevado a cabo durante estos últimos años en este campo se refiere al estudio de las reacciones inducidas o catalizadas mediante el láser a través de un proceso de excitación multifotónica (o absorción consecutiva de varios fotones) producido por un láser de infrarrojos. Entre las diferentes razones que justifican la realización de dichos estudios cabe destacar fundamentalmente dos. La primera de ellas se refiere a la disponibilidad de láseres de infrarrojos de alta potencia y fácil regulación. El láser de dióxido de carbono, que es el caballo de batalla de la industria, es capaz de suministrar un mol de fotones ($6{,}02 \times 10^{23}$ fotones) a un coste total de tres centavos de dólar. La segunda razón estriba en la necesidad de investigar el papel que desempeñan los estados excitados de vibración de alta energía en la realización de las reacciones químicas, una incógnita que no se ha podido desvelar durante varias décadas.

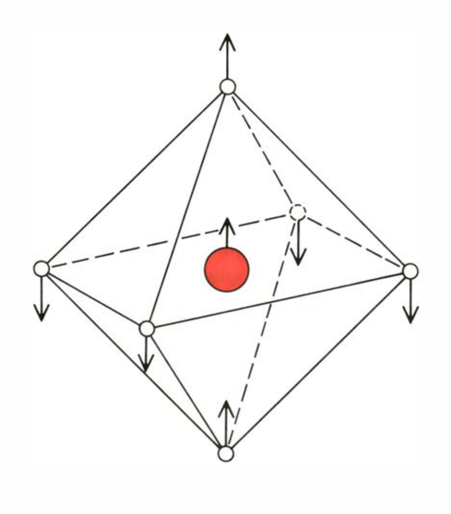
Vamos a referirnos ahora a dos reacciones químicas excitadas por láser que están bien estudiadas. La primera de estas reacciones consiste en la disociación del hexafluoruro de azufre (SF_6), que es quizás el ejemplo mejor estudiado de una disociación unimolecular inducida por el láser y cuyo origen radica en la absorción múltiple de fotones infrarrojos. El segundo ejemplo se refiere a una reacción bimolecular catalizada por el láser en la cual el monóxido de nitrógeno (NO) reacciona con el ozono (O_3) para formar dióxido de nitrógeno (NO_2) y oxígeno (O_2). Esta última reacción se cataliza mediante un único fotón infrarrojo cuya energía se ha ajustado de forma tal que se produzca su absorción por una molécula de NO o por una molécula de O_3.

La disociación del SF_6 transcurre como sigue. Los primeros tres o cuatro fotones infrarrojos resultan absorbidos casi simultáneamente y para una misma frecuencia, provocando la transición de la molécula de SF_6 desde el estado fundamental hasta estados excitados de vibración de la molécula de energía cada vez mayor. Por encima del último de estos estados, el número de niveles energéticos disponibles resulta ser extraordinariamente elevado (más de 10^9 estados por electronvolt). Esta circunstancia reduce sustancialmente las restricciones inherentes a la longitud de onda de la radiación del láser incidente. En esta región, conocida como la zona del casi continuo, los estados adicionales de rotación y traslación proporcionan los necesarios "ajustes finos" para igualar el intervalo existente entre los niveles vibracionales con la frecuencia de los fotones. En otras palabras, las condiciones de resonancia exacta entre la frecuencia de la radiación y la energía de los estados excitados se han de cumplir tan sólo para los tres o cuatro primeros fotones; los fotones subsiguientes a éstos resultarán absorbidos aun cuando la distribución de sus longitudes de onda sea mucho más amplia. En conjunto, su absorción conducirá a la molécula en cuestión hasta el límite de disociación. El número



SEPARACION DE ISOTOPOS. Puede conseguirse en algunos casos procediendo a disociar de forma preferente aquellas moléculas que incluyan un isótopo determinado de cierto elemento sin que se produzca perturbación alguna sobre las moléculas que, siendo similares a las anteriores, incluyen un isótopo diferente. Procediendo de este modo, se puede separar el ^{34}S, un isótopo raro del azufre natural, del ^{32}S, mucho más abundante. En el hexafluoruro de azufre ordinario existe una mezcla de estos dos isótopos ($^{34}SF_6$ y $^{32}SF_6$). Un láser de infrarrojos, ajustado exactamente a 10,61 micrometros, producirá la excitación de los estados vibracionales del $^{32}SF_6$, pero no los del $^{34}SF_6$. Las flechas representan la absorción sucesiva de fotones. Cuando se alcanza la región del continuo el $^{32}SF_6$ se disociará en $^{32}SF_5$ y flúor.





VIBRACIONES DEL HEXAFLUORURO DE AZUFRE excitadas mediante el empleo de radiación infrarroja. Estas moléculas están constituidas por un átomo de azufre (*color*) rodeado por seis átomos de flúor (*círculos*), que se disponen en los vértices de un octaedro. Para mayor claridad, los enlaces azufre-flúor no se han representado en esta figura. En uno de los modos fundamentales de vibración (*izquierda*), el átomo de azufre se mueve acercándose y alejándose a dos átomos de flúor que están separados entre sí 180 grados. En el segundo modo (*derecha*), el átomo de azufre y los dos de flúor situados en un eje común se mueven juntos, y los otros átomos de flúor se mueven en sentido contrario.

total de fotones infrarrojos necesarios para dicho fin es igual a 35. Este número corresponde a la energía necesaria para romper el enlace simple entre el azufre y el flúor (80 kilocalorías por mol).

Se ha demostrado que la disociación comienza cuando se produce la ruptura de un enlace simple azufre-flúor y se forma el radical SF_5 y un átomo de flúor. La disociación es tan rápida que no hay tiempo para que la energía se pueda transferir por colisiones de una molécula a otra. Sólo se romperán aquellas moléculas que hayan absorbido fotones del láser. En este proceso no se forma ningún tipo de subproductos; al parecer, la disociación del SF_6 constituye un caso ideal de excitación específica de un enlace y disociación específica de ese mismo enlace.

La eficaz reacción de disociación del SF_6 ha puesto de manifiesto que la radiación del láser se puede utilizar también para separar isótopos; en este caso se trata de las dos especies nucleares del azufre: el azufre 32 (escrito ^{32}S) y el azufre 34 (^{34}S). Este último isótopo posee dos neutrones más que el primero y constituye tan sólo el 4 por ciento de los átomos de azufre de la naturaleza, mientras que el 96 por ciento restante corresponde al isótopo ^{32}S. El objetivo de este experimento consiste en excitar y descomponer selectivamente las moléculas de $^{32}SF_6$, con lo cual el gas residual se enriquece en $^{34}SF_6$ (o viceversa). La primera separación que se vio coronada por el éxito fue llevada a cabo por R. V. Ambartzumian y V. S. Letokhov, del Instituto de Espectroscopía de Moscú. Estos autores procedieron a irradiar hexafluoruro de azufre con un láser pulsante de dióxido de carbono ajustado de forma tal que su frecuencia coincidiera con la frecuencia de absorción vibracional del $^{32}SF_6$. Virtualmente todo el $^{32}SF_6$ se descompuso, dejando un gas residual enriquecido en $^{34}SF_6$ casi 3000 veces más que el original. La frecuencia efectiva del láser fue de 10,61 micrometros. Cuando la frecuencia se desplazaba ligeramente hasta el valor de 10,82 micrometros, las moléculas de $^{34}SF_6$ se descomponían selectivamente. Este método de separación de isótopos está siendo estudiado intensamente en cuanto se refiere a su aplicación a la separación del uranio 235 (fisionable) del uranio 238 (no fisionable). En relación con este tema puede consultarse el artículo titulado "La separación de isótopos con láser" de Richard N. Zare, que fue publicado en INVESTIGACION Y CIENCIA en abril de 1977.

La descomposición del SF_6 plantea una cuestión importante para la química del láser: ¿representa esta descomposición realmente la excitación específica de un enlace seguida por una disociación específica de dicho enlace? Si tomamos en consideración el breve tiempo que transcurre entre la absorción de los fotones del láser y la disociación de la molécula, la respuesta debería ser afirmativa.

Sin embargo, el hecho de que esta molécula posea seis átomos de flúor químicamente equivalentes plantea una duda. ¿Cómo podemos asegurar que el enlace que se rompe es precisamente aquel que ha absorbido todos los fotones del láser o, al menos, la mayor parte de ellos? Para contestar a esta pregunta se podría plantear, en principio, la realización de un experimento decisivo, consistente en reemplazar uno de los átomos de flúor de esta molécula por un átomo de otro isótopo de dicho elemento. Sin embargo, la realización de este experimento es posible en la práctica, dado que sólo existe un isótopo estable del flúor (^{19}F) y que el isótopo radiactivo de vida más larga con que cuenta este elemento (^{18}F) tiene un período de semidesintegración inferior a las dos horas.

Afortunadamente, un experimento alternativo nos permite resolver este problema. Uno de los átomos de flúor se puede reemplazar por otro de cloro para dar lugar al compuesto SF_5Cl. Se repite entonces el experimento con el láser ajustado para excitar la misma vibración del azufre-flúor que la utilizada para el SF_6. Inesperadamente, se encuentra que la citada molécula se descompone en SF_5 y Cl en lugar de liberar un átomo de flúor. La explicación aparente de este resultado se basa en el hecho de que uno de los enlaces de azufre-flúor absorbe la energía del láser y que las vibraciones son rápidamente compartidas (sin que tengan lugar colisiones) por todos los átomos ligados al azufre central. Cuando el cloro es uno de estos átomos, su enlace con el azufre, más débil que los otros, es precisamente el enlace que se rompe. Evidentemente, cuando los seis átomos son de flúor, el enlace que se rompe es una cuestión de azar; todos los enlaces son igualmente fuertes. La conclusión de todo este razonamiento es que la disociación del SF_6 no tiene realmente un carácter específico respecto al enlace. En otros muchos experimentos de absorción múltiple de fotones infrarojos con otras especies moleculares se ha llegado a conclusiones similares a la anterior. La implicación de tales estudios en relación con la química del láser supone que cuando se prepara un estado transitorio por absorción múltiple de fotones, lo más probable es que se rompa aquel enlace que sea más débil.

Esta afirmación es correcta sólo en el caso de los estudios llevados a cabo en la actualidad utilizando impulsos de láser relativamente largos (de 0,01 microsegundos a 3 microsegundos). Se predice teóricamente que al utilizar impulsos unas 10.000 veces más cortos (de 1 picosegundo a 100 picosegundos) sí que se conseguirá una excitación y ruptura de enlace selectiva. La razón de esta diferencia estriba en el hecho de que se necesita un cierto tiempo mínimo, cuyo valor se estima que es del orden de 100 picosegundos, para que la energía se reparta en el interior de una molécula excitada. En consecuencia, si se utilizan

impulsos de láser cuya duración sea menor que 100 picosegundos, se producirá la disociación del enlace individual de la molécula considerada antes de que la energía absorbida por ésta se reparta entre todos los enlaces de la misma.

A continuación vamos a estudiar con más detenimiento la reacción bimolecular catalizada por la absorción de un único fotón del láser a la que hicimos referencia anteriormente: la reacción entre el monóxido de nitrógeno y el ozono. Esta reacción entre una molécula diatómica y una molécula triatómica tiene lugar por excitación infrarroja selectiva de una vibración determinada en cada una de estas dos moléculas. El ozono se excita selectivamente mediante un láser de dióxido de carbono con una longitud de onda de 9,48 micrometros. El monóxido de nitrógeno se excita por separado mediante un láser de monóxido de carbono con una longitud de onda de 5,2 micrometros. La reacción en cuestión requiere una energía de excitación pequeña: unas cinco kilocalorías por mol de reactivos.

Dado que esta reacción transcurre a velocidad apreciable incluso a temperatura ambiente, cabe esperar que la deposición de unas 2,3 kilocalorías de energía radiante en un estado excitado específico del ozono mediante un láser de dióxido de carbono producirá un aumento importante en la velocidad de reacción. El aumento en dicha velocidad sería todavía mayor si se produjera la deposición de unas 4 kilocalorías en el modo de vibración del monóxido de nitrógeno mediante un láser de monóxido de carbono. En cada uno de estos experimentos, la energía suministrada corresponde aproximadamente a un fotón por molécula. En ambos experimentos la velocidad de reacción aumentaba en un factor de 20 sobre el valor observado a la temperatura ambiente.

La intensificación selectiva de una reacción parece ser una consecuencia de la excitación específica de un enlace producida por el láser en uno u otro de los dos reactivos citados y transferida a los productos de la reacción a través de un proceso intermedio. Sin embargo, incluso en este caso, el mecanismo real es más complejo de lo que cabría esperar. Los estudios de transferencia de energía han puesto de manifiesto que las colisiones desempeñan un papel importante en esta reacción catalizada por una radiación infrarroja. En el caso de la molécula de ozono, se puede activar por colisión más de un estado de vibración, como proceso subsiguiente a la excitación específica de una vibración con ayuda del láser. Esta activación tiene lugar en el transcurso de la reacción considerada. En consecuencia, al menos dos, y posiblemente más, de las coordenadas de la reacción correspondientes a las vibraciones excitadas son responsables de la intensificación catalítica observada. Por otra parte, es difícil asegurar si la excitación es estrictamente vibracional o si, como resultado de la producción de colisiones, la excitación traslacional contribuye también de forma significativa a este proceso. Las experiencias más recientes que se han llevado a cabo en este sentido parecen indicar que las vibraciones y las traslaciones tienen la misma importancia en la intensificación de la velocidad de una reacción.

Aunque estos resultados suponen la superación de algunos de los problemas que empañaban el brillante futuro de la aplicación del láser a la tecnología química, también pueden ser considerados como un frenazo importante ante un entusiasmo prematuro. Como es natural, los primeros resultados publicados sobre este tema no se muestran exentos de ciertas interpretaciones erróneas. La química del láser es un campo de trabajo cuyo conocimiento está teniendo lugar a un ritmo explosivo; la experiencia indica que la comprensión más profunda de un tema surge con frecuencia a la luz de una reinterpretación de los resultados preliminares.

Los anteriores ejemplos han puesto de manifiesto con claridad algunos de los problemas que caracterizan a las reacciones químicas que transcurren en una escala temporal del orden del microsegundo, en cuyos procesos intervienen fenómenos tan variados como las vibraciones de los enlaces, las colisiones moleculares y la distribución de energía en el seno de las moléculas que sufren una disociación con o sin que tengan lugar choques con otras moléculas. Sin embargo, la síntesis de especies químicas con ayuda del láser, producida en volúmenes muy pequeños, es ya una realidad. En la actualidad se han conseguido preparar ciertos compuestos que, a nivel de los métodos usuales de detección, parecen estar totalmente exentos de subproductos. Algunos productos, difíciles de aislar por métodos convencionales, han sido separados con éxito mediante la utilización del láser.

El principal obstáculo que presenta la introducción a gran escala del láser en la industria química es de orden económico. Hoy por hoy, los fotones del láser son considerablemente más caros que los que llegan a una vasija de reacción procedentes de una corriente de vapor o de la llama que se produce al quemar un combustible fósil. El coste de un mol de fotones térmicos es infinitésimo. El coste de un mol de fotones de láser es del or-

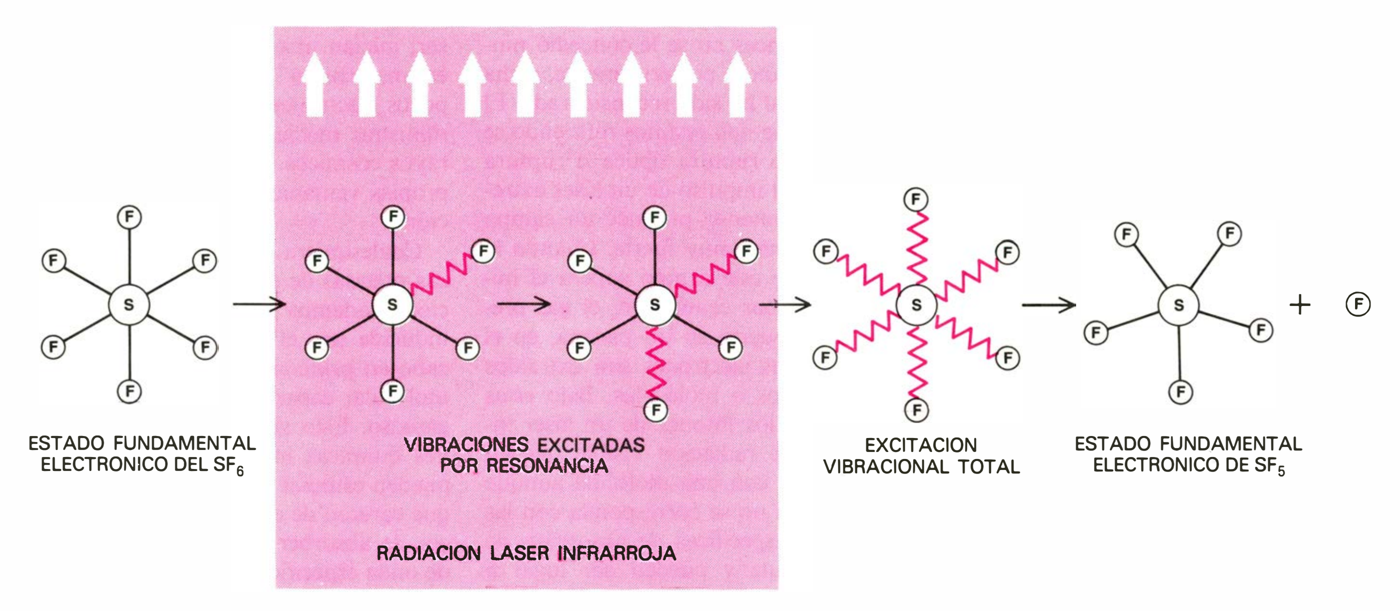
den de algunos centavos de dólar para fotones con longitudes de onda en el infrarrojo y de algunos dólares para fotones con longitudes de onda en la región visible o ultravioleta. Los láseres comerciales de dióxido de carbono operan en el infrarrojo con una eficiencia del orden de 5×10^{19} fotones por cada impulso de un microsegundo (este valor equivale a la energía de un joule); en consecuencia, se necesitan unos 10.000 impulsos para proporcionar un mol de fotones. Si la reacción se produce gracias a la acción de un láser, en términos de mol-por-mol, el coste energético en números redondos será del orden de algunos centavos por varias decenas de gramos de producto. Evidentemente, las primeras aplicaciones de este método consistirán en la preparación de ciertos productos especiales –como productos químicos especiales, farmacéuticos, catalizadores, isótopos raros– cuyo coste supera ampliamente el valor de algunos dólares por libra que resulta ser el usual en el caso de los productos químicos más corrientes.

Por otra parte, cuando el láser se utiliza como catalizador, en cuyo papel da lugar a muchos moles de producto por cada mol de fotones, el aspecto económico resulta todavía mucho más atractivo. La cuestión radica en encontrar reacciones que se puedan producir con relativa facilidad por encima de la energía umbral necesaria para la formación de un estado excitado que a continuación inicie una reacción en cadena. Estas reacciones se utilizan en la preparación de muchos productos químicos industriales y sólo requieren una pequeña cantidad de catalizador por mol de producto. Quizás el uso más conocido de los catalizadores resida en la industria petroquímica, en la que la presencia de tales sustancias ayuda a romper las grandes moléculas para obtener otras más pequeñas, a polimerizar las pequeñas para obtener otras mayores y a recomponer las estructuras moleculares en configuraciones más adecuadas para algún fin determinado.

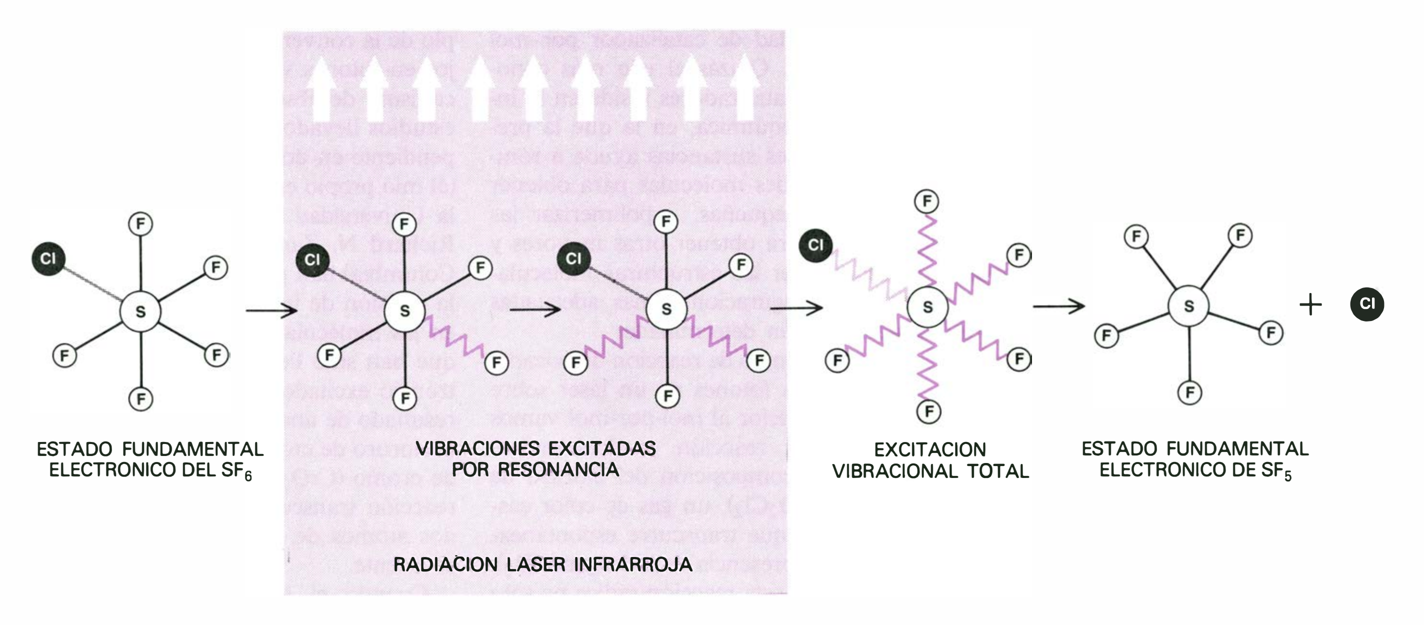
Como ejemplo de reacción desencadenada por los fotones de un láser sobre una base inferior al mol-por-mol vamos a citar una reacción particularmente rara: la descomposición del cloruro de cromilo (CrO_2Cl_2), un gas de color castaño rojizo, que transcurre espontáneamente o en presencia de hidrógeno (H_2). El interés de esta reacción radica no sólo en el uso económico de fotones que la caracteriza, sino también en el hecho de que el cromo presente en el producto resultante de la misma se presenta combinado con oxígeno en diferentes proporciones, dependiendo de la presencia o la ausencia de un eliminador químico, que en este caso es el hidrógeno al que anteriormente nos hemos referido.

En la primera versión de esta experiencia, con el hidrógeno ausente, se procedió a irradiar el cloruro de cromilo gaseoso a baja presión mediante un láser de dióxido de carbono, cuya longitud de onda es de 10 micrometros. Como resultado de la absorción de fotones infrarrojos, el cloruro de cromilo emite un espectro amplio de luz visible, con predominio en el color naranja. Esta reacción constituye el primer ejemplo de la conversión de fotones infrarrojos en fotones visibles mediante un mecanismo de absorción y reemisión. Los estudios llevados a cabo de forma independiente en dos laboratorios diferentes (el mío propio en el Brooklyn College de la Universidad de Nueva York y el de Richard N. Zare en la Universidad de Columbia) han puesto de manifiesto que la emisión de luz visible tiene su origen en las moléculas de cloruro de cromilo que han sido llevadas a un estado electrónico excitado de baja energía. Como resultado de una excitación más intensa, el cloruro de cromilo se disocia en óxido de cromo (CrO_2) y gas cloro (Cl_2). Esta reacción transcurre en dos etapas y los dos átomos de cloro se liberan secuencialmente.

Cuando el hidrógeno no está presente, la emisión de luz visible coincide exactamente con la duración del impulso del láser. Como consecuencia de este hecho, cada molécula absorbente se eleva directamente, sin que se produzcan colisiones moleculares, desde un nivel de energía vibracional situado en el estado fundamental hasta un nivel de energía vibracional situado en el estado excitado electrónico. Para alcanzar el estado excitado de energía más elevada, una molécula de cloruro de cromilo debe absorber de 16 a 18 fotones infrarrojos en rápida sucesión. Al llegar a este punto, la molécula puede liberar un solo fotón visible de alta energía y retornar a su estado fundamental original o bien absorber 10 fotones adicionales y utilizar toda la energía que ha absorbido para disociarse en CrO_2Cl y un átomo de cloro.



DISOCIACION DEL HEXAFLUORURO DE AZUFRE, representada esquemáticamente en esta figura. Se produce por la absorción secuencial de 35 fotones infrarrojos procedentes de un láser de dióxido de carbono. La energía transportada por los fotones se transfiere a los enlaces de azufre-flúor que constituyen dicha molécula (*color*). Como consecuencia de dicho proceso, la molécula en cuestión sufre una serie de transiciones sucesivas a diferentes estados vibracionales de alta energía. Finalmente, se rompe uno de los seis enlaces de azufre-flúor y queda en libertad un átomo de flúor. El fragmento molecular restante, pentafluoruro de azufre (SF_5), retiene cualquier exceso de energía que exista tras la ruptura del enlace anterior.



SUSTITUCION DE UN ATOMO DE FLUOR por un átomo de cloro en el hexafluoruro de azufre. Tal sustitución conduce al SF_5Cl. Este compuesto se puede utilizar para dilucidar el mecanismo según el cual se produce la disociación de dicha molécula mediante el láser. Si el láser en cuestión se ajusta de forma tal que excite específicamente los enlaces de azufre-flúor, cabe esperar que se rompa uno de estos enlaces. Ahora bien, lo que en realidad se rompe es el enlace de azufre-cloro, con lo cual queda en libertad un átomo de cloro. Aunque la energía del fotón se absorbe inicialmente por los enlaces azufre-flúor, la excitación vibracional se reparte rápidamente entre todos los enlaces de la molécula. Dado que el enlace de azufre-cloro es más débil que cualquiera de los enlaces de azufre-flúor, el enlace que se rompe en esta molécula es el primero de los citados.

Dada la duración de su existencia, el fragmento de CrO_2Cl está también sujeto a la excitación con el láser. Este fragmento absorbe de 10 a 16 fotones y, a su vez, emite radiación visible o utiliza la energía absorbida para disociarse en CrO_2 y otro átomo de cloro. En consecuencia, los productos finales son el CrO_2 (un polvo de color marrón) y el gas cloro (Cl_2).

Cuando se añaden 1,5 moles de hidrógeno por cada mol de cloruro de cromilo, el camino que sigue la reacción es inicialmente el mismo que antes: la liberación de un átomo de cloro a partir de una pequeña fracción de las moléculas de cloruro de cromilo. A continuación, la reacción transcurre ya en una forma muy diferente. El átomo de cloro que se ha disociado reacciona con el hidrógeno molecular en una reacción en cadena bien conocida que desdobla a la molécula de H_2 en los dos átomos de hidrógeno que la constituyen y forma cloruro de hidrógeno (HCl). El otro átomo de hidrógeno reacciona con el CrO_2Cl parcialmente disociado formando más HCl y agua junto con otras especies hidroxiladas más complejas. Los productos finales son HCl, agua y una forma diferente de óxido de cromo (Cr_2O_3).

En ambas reacciones el cromo estaba originalmente en su sexto estado de oxidación, lo cual supone que para llevarlo a un estado neutro es necesaria la incorporación de seis electrones a cada átomo. Si la reacción tiene lugar sin la presencia de hidrógeno, el cromo se convierte al cuarto estado de oxidación. Si la reacción se produce en presencia de hidrógeno, el cromo pasa al tercer estado de oxidación. En esta segunda reacción tan sólo se necesita una fracción de un mol de fotones para convertir a un mol de cloruro de cromilo en un mol de óxido de cromo. La disociación del cloruro de cromilo en presencia de hidrógeno es una de las pocas reacciones en cadena con ramificaciones que se han estudiado utilizando la técnica del láser.

Todavía existe otro mecanismo mediante el cual las reacciones químicas se pueden activar con un láser. El fenómeno en cuestión ya se discutió cuando se empezó a utilizar este instrumento, aunque entonces no se le concedió ninguna importancia; posteriormente, dicha postura inicial ha sido reconsiderada. El proceso al que nos estamos refiriendo se conoce como ruptura óptica o ruptura dieléctrica. El impulso de un láser extremadamente intenso produce un campo eléctrico alterno muy fuerte. Cuando la intensidad de este campo supera el millón de volt por centímetro, el gas presente se convierte en un plasma, en el cual todos los electrones son extraídos de sus átomos o moléculas. Bajo estas condiciones, los fotones de un láser infrarrojo o de radiación visible pueden interaccionar con una molécula aunque su frecuencia no se corresponda con las frecuencias específicas de absorción de dicha molécula y pueden dar lugar a una reacción relativamente específica a través del fenómeno de ruptura. Las reacciones producidas por la acción de dichos campos eléctricos intensos, que interaccionan con la molécula como un todo, son similares a las reacciones químicas que se producen mediante temperaturas muy elevadas (por ejemplo, utilizando una descarga eléctrica) y que tienen lugar en una escala de tiempo del orden del microsegundo.

Los resultados encontrados en diferentes estudios han demostrado que la ruptura óptica que se produce en el campo intenso de un láser de dióxido de carbono consiste en una reacción en cadena conocida como ionización en avalancha. Este mecanismo exige la presencia de un pequeño número inicial de electrones. Cuando dichos electrones sufren una aceleración en el campo del láser, inician una avalancha que culmina en una "chispa" de plasma ionizado. Los pocos electrones iniciales se pueden suministrar mediante partículas de polvo, rayos cósmicos o incluso a través de las propias ventanas del recipiente de reacción.

Cualesquiera que sean las características precisas de este mecanismo de reacción, podemos afirmar que la ruptura inducida por el láser se puede llevar a cabo en principio con cualquier especie molecular capaz de ser llevada al estado gaseoso. Esto significa que las reacciones químicas inducidas por el láser se pueden estudiar para muchas sustancias que carecen de estados energéticos capaces de absorber fotones con longitudes de onda específicas disponibles en los láseres. Evidentemente, se hace necesario proceder a un estudio caso por caso con el fin de determinar si la ruptura inducida por el láser resulta ventajosa o no respecto a la aplicación directa de calor

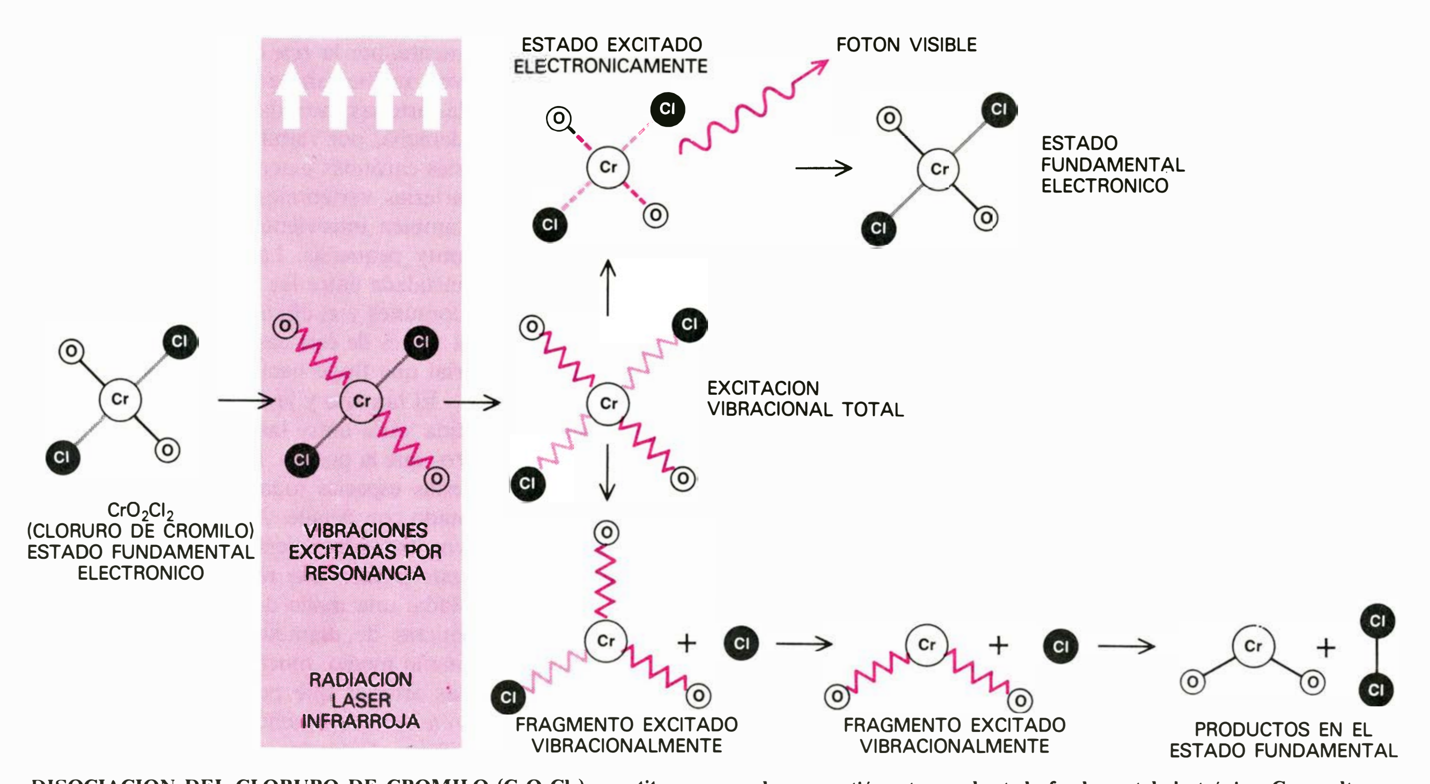
o de simples descargas eléctricas continuas. Sin embargo, ciertos estudios han demostrado que la ruptura inducida por el láser es capaz de proporcionar ciertos productos específicos que resultan más difíciles de obtener recurriendo a otros métodos de excitación.

Por otra parte, las reacciones inducidas por una ruptura dieléctrica producida por un láser pueden ofrecer ciertas ventajas económicas en relación con las reacciones de excitación específicas. La eficiencia de las reacciones de este último tipo resulta ser máxima cuando los reactivos gaseosos están sometidos a baja presión. Este hecho supone una seria limitación en relación con el volumen del producto que se puede obtener utilizando una cámara de reacción de un tamaño determinado. Por el contrario, la ruptura dieléctrica inducida por el láser permite que los gases reaccionantes se encuentren a presiones muy elevadas y, en consecuencia, el rendimiento de la correspondiente reacción resulta ser considerablemente mayor. Si a este hecho se añade la posibilidad de que la ruptura dieléctrica pueda conducir también a una reacción en cadena, las perspectivas económicas resultan ser todavía más atractivas.

En mi propio laboratorio se han llevado a cabo una serie de estudios sobre la producción de ciertas reacciones por absorción múltiple de fotones y por ruptura dieléctrica inducida por láser, como resultado de las cuales se obtuvieron productos idénticos. Este hecho no significa necesariamente que las reacciones en cuestión transcurran según los mismos mecanismos en ambos casos, sino simplemente que las condiciones termodinámicas son tales que las reacciones tienen lugar a través de los canales disponibles cuya energía sea menor y que conducen al mismo resultado final. De hecho, se puede plantear la duda de si los mecanismos de reacción son realmente idénticos, puesto que hay pruebas en el sentido de que los procesos de absorción múltiple de fotones resultan controlados por las mismas restricciones que rigen los procesos inducidos por la ruptura dieléctrica. Por ejemplo, la existencia de determinados radicales o fragmentos moleculares, tales como el C_2, CH, NH_2, CrO_2Cl y CrO_2, se observa tanto en aquellas reacciones que son inducidas por la absorción multifotónica como por la ruptura dieléctrica. Por otra parte, ambos tipos de experiencias se pueden interpretar teóricamente mediante la aplicación de las mismas ecuaciones termodinámicas y estadísticas.

La ruptura óptica inducida por el láser se puede aplicar también a la preparación de sólidos finamente divididos. Por ejemplo, cuando el sulfuro de carbonilo (OCS) se descompone mediante este mecanismo, en la cámara de reacción precipita azufre muy fino [*véase la ilustración de la portada de esta revista*]. La precipitación del azufre se puede controlar con precisión a través de la presión del gas, las condiciones operativas del láser y la velocidad de extinción. Estos mismos principios parecen ser aplicables también a la preparación de otras sustancias en un estado de división fina.

Varias ramas de la industria química están estudiando con gran atención las aplicaciones del láser a los procesos comerciales. Los éxitos logrados a escala de laboratorio mediante el láser en la separación de isótopos, la separación química, la síntesis química y la extracción de sustancias traza de corrientes gaseosas a las que se exige un grado de pureza muy elevado, demandan la realización de un esfuerzo muy considerable en el aspecto de desarrollo. El motivo por el cual los láseres no han producido un impacto todavía mayor en la industria química es puramente económico. Sin embargo, la tecnología del láser sigue progresando de forma continua y el coste de los dispositivos utilizados disminuye en términos relativos a medida que aumenta su eficiencia. En la actualidad los costes serían ya competitivos en el caso de productos farmacéuticos y otros productos químicos delicados. Es muy probable que en los próximos años se descubran nuevas aplicaciones para el láser.



DISOCIACION DEL CLORURO DE CROMILO (CrO_2Cl_2); constituye un caso peculiar entre las reacciones inducidas por láser. Una vez que esta molécula ha sido llevada hasta un estado transitorio de excitación vibracional gracias a la absorción de 16 a 18 fotones infrarrojos, puede desprenderse de toda la energía absorbida emitiendo un solo fotón de luz visible, cuya energía es mucho mayor que la de los anteriores, con lo cual la molécula en cuestión retorna al estado fundamental electrónico. Como alternativa, si se produce la absorción de otros 10 fotones, la molécula excitada se puede descomponer a través de un proceso en dos etapas, dando lugar a óxido de cromo (CrO_2) y una molécula de cloro (Cl_2). La disociación del cloruro de cromilo y de una mezcla de cloruro de cromilo e hidrógeno, provocada con la ayuda del láser, se muestra en las fotografías precedentes.

Un sistema de enfriamiento del cerebro en los mamíferos

Los mamíferos carnívoros y algunas de sus presas pertenecientes a la misma clase toleran extremos letales de calor y ejercicio debido a que tienen una rete, *o red intercambiadora de calor, que evita que el cerebro se caliente demasiado*

Mary Ann Baker

En un día caluroso un perro puede perseguir a un conejo hasta la muerte, por asfixia, de éste. ¿Cómo es ello posible? Porque, si bien la carrera eleva la temperatura de ambos, el cerebro del perro tiene un sistema de enfriamiento del que carece el cerebro del conejo. Si el conejo no consigue dar con una madriguera donde esconderse y enfriarse, la temperatura de su cerebro pronto alcanzará un nivel letal.

En concreto, lo que el perro tiene, y no el conejo, es un plexo intercambiador de calor por contracorriente en la base de su cerebro, estructura anatómica hallada en algunos mamíferos aunque no en todos. Este plexo consiste en una red de pequeños vasos sanguíneos que se ramifican desde las arterias carótidas. Fue descrito por primera vez hace más de dos milenios por el médico griego Herófilo. Cuando estaba diseccionando la cabeza de un animal (probablemente una oveja) aproximadamente hacia el 300 a. C. descubrió una conspicua red de vasos sanguíneos en la base del cerebro.

Aunque Herófilo describió ya la estructura, fue el anatomista Galeno quien casi medio milenio más tarde la hizo famosa. Trabajando a la sazón en Roma, a mediados del segundo siglo de nuestra era, asignó las funciones que competían a la red arterial y les acreditaba como responsables del término que pronto recibió: *rete mirabile*, o red maravillosa. Galeno no pudo conseguir cadáveres humanos para disección; su anatomía descriptiva se basa fundamentalmente en el estudio de animales domésticos.

Desde su punto de vista, la *rete* carótida era una estructura anatómica clave; actuaba transformando el "espíritu vital", que transportaban las arterias, en el "espíritu psíquico". Este espíritu era transportado luego a través del cuerpo por el sistema nervioso, del que Galeno suponía que estaba formado por tubos huecos.

Galeno ejerció un profunda influencia en la anatomía y medicina posterior. Desde su tiempo hasta el Renacimiento, los anatomistas que poseían cabezas humanas para diseccionar confiaban en encontrar la *rete* maravillosa apoyados en la autoridad del maestro Galeno, que la había descrito, a pesar de que no se hallara ni en el hombre ni en los restantes primates. El gran anatomista del Renacimiento Vesalio, en sus escritos de 1538, expone el siguiente dilema: "Yo que tanto he trabajado siguiendo los pasos de Galeno... nunca he emprendido la disección de una cabeza humana en público sin tener al lado la de un cordero o buey, con el objeto de presentar lo que no podía encontrar de ninguna manera en el hombre, resaltándola ante los espectadores, por miedo a que se me acusara de incapacidad de hallar el plexo cuyo nombre es tan universalmente conocido".

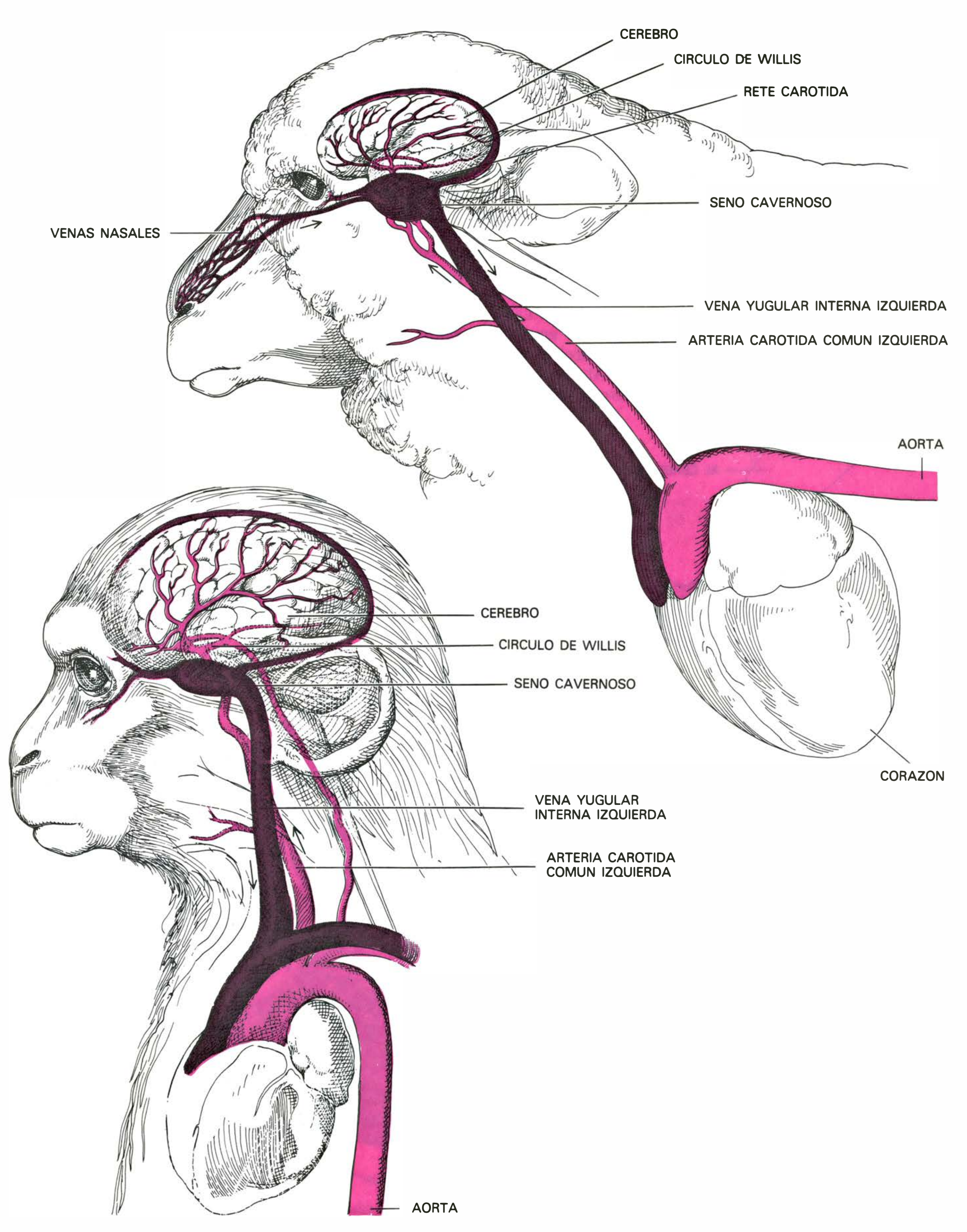
Las redes de intercambio de calor por contracorriente no están limitadas al cerebro de ciertos mamíferos; se localizan en otras zonas del cuerpo en la escala animal. Son las responsables de que las aves zancudas puedan aguantar largos períodos en agua fría, de que algunos mamíferos marinos y peces puedan vivir en mares polares y de que mamíferos tropicales como el perezoso y el oso hormiguero puedan conservar el calor corporal durante la noche. Pero aquí vamos a tratar sólo de la primera red maravillosa descubierta, la *rete* carótida, y el papel que desempeña en el mantenimiento del enfriamiento del cerebro.

Falta la *rete* carótida en los monotremas (orden en que se encuadra el ornitorrinco), los marsupiales, los perisodáctilos (orden al que pertenecen el caballo, el rinoceronte y el tapir), los roedores, los lagomorfos (orden que incluye conejos y liebres) y en los primates. Así, el cerebro humano es abastecido de sangre por medio de las arterias carótidas internas derecha e izquierda y por las arterias vertebrales derecha e izquierda. En la base del cerebro las cuatro arterias se conectan en el círculo de Willis, en honor del anatomista inglés que lo describió en el siglo XVII. A partir de este círculo, surgen todas las grandes arterias que nutren de sangre al cerebro.

Entre los artiodáctilos (orden que incluye al ganado vacuno y a otros mamíferos de pezuña con número par de dedos) y muchos carnívoros, el modelo de irrigación sanguínea cerebral difiere bastante. En vez de recibir la sangre de los dos pares de arterias craneales, el círculo de Willis en estos animales queda abastecido, fundamental o totalmente, por la *rete* carótida. La *rete*, a su vez, recibe sangre ya sea por ramas de las arterias carótidas internas izquierda y derecha, por ramificaciones de las arterias carótidas externas, o de ambas. Las arterias vertebrales derecha e izquierda también intervienen, aunque suelen ser muy pequeñas. La *rete* se halla, pues, instalada entre las dos arterias carótidas comunes y el círculo de Willis, pasando a través de éste casi toda la sangre arterial que fluye hacia el cerebro.

El tamaño y la forma de la *rete* carótida varía entre las especies de mamíferos que la poseen, aunque en algunas de estas especies todavía no se ha examinado con detalle. Así, entre los carnívoros, todos los miembros de la familia del gato poseen una red muy bien desarrollada: una malla de vasos de 200 a 300 micras de diámetro. Son vasos de tamaño medio, mucho más pequeños que las arterias que penetran y salen de la *rete* aunque mucho mayores que los capilares. Entre los perros, lobos e hienas y entre algunas focas y leones marinos la red es pequeña y consta sólo de arterias de tamaño medio entrelazadas.

Todos los mamíferos, los que portan *rete* carótida y los que carecen de ella.

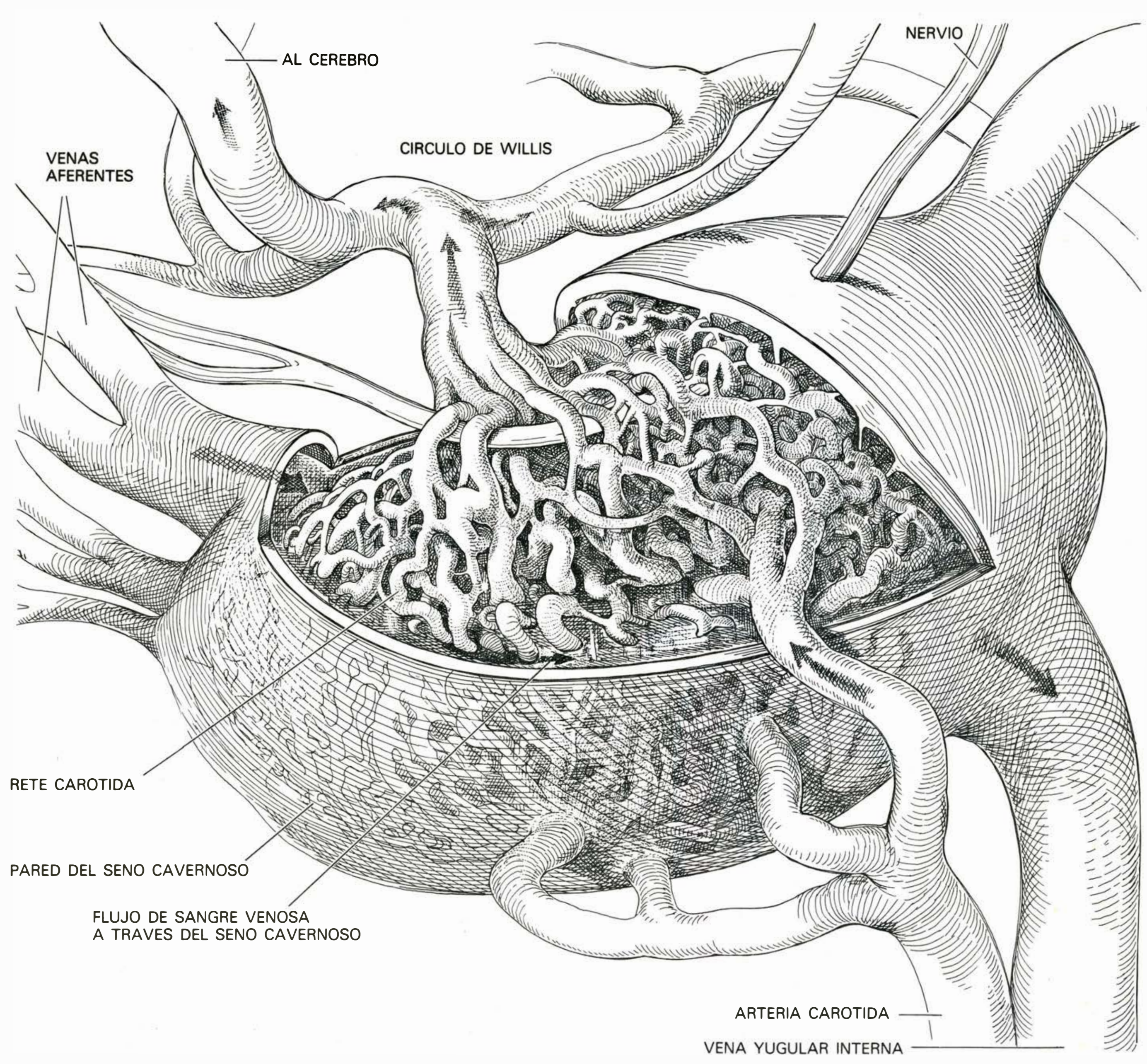


APORTE DE SANGRE al cerebro de un mamífero. Desde el corazón, el flujo sigue uno de dos caminos fundamentales según tenga o no *rete* carótida. Como se ilustra esquemáticamente aquí, la oveja (*arriba*) posee *rete* pero no el mono (*abajo*). En la oveja la sangre arterial irriga el cerebro a partir de una red de vasos conocida por círculo de Willis. Parte de la sangre llega al círculo por las arterias vertebrales, pero el caudal mayor procede de las arterias carótidas comunes y sus ramas, dando lugar a una malla de vasos que forman la *rete* carótida. Esta red, situada en el interior de un depósito de sangre venosa en una cavidad denominada seno cavernoso, actúa como intercambiador de calor; la sangre arterial más caliente pierde calor en favor de la sangre venosa, más fría. De ello resulta que la mayor parte de la sangre arterial que alcanza el cerebro de la oveja está a una temperatura menor que la que tenía cuando fue expulsada del corazón. En el mono, la sangre arterial viaja desde el corazón, a través de las arterias vertebral carótida y sus ramas, directamente hasta el círculo de Willis. Las arterias carótidas internas pasan a través del seno cavernoso con una pérdida de calor mínima hacia la sangre venosa, de modo que la sangre irriga el cerebro a la temperatura del corazón. (La *rete* falta en el hombre.)

muestran la misma disposición anatómica en la base del cerebro. Aquí, que es por donde las arterias que suministran sangre al cerebro entran en la cavidad craneana, se encuentran grandes reservorios de sangre venosa. Conocidos como senos venosos, los reservorios reciben sangre de las venas procedentes del exterior y del interior del cráneo. Uno de esos reservorios recibe la denominación de seno cavernoso. En los mamíferos sin *rete* las arterias carótidas internas atraviesan el seno cavernoso en su camino hacia el círculo de Willis. En los animales dotados de *rete*, el plexo arterial se dispone en el interior del seno cavernoso o está asociado con un plexo similar de vasos de sangre venosa conectado con dicho seno cavernoso.

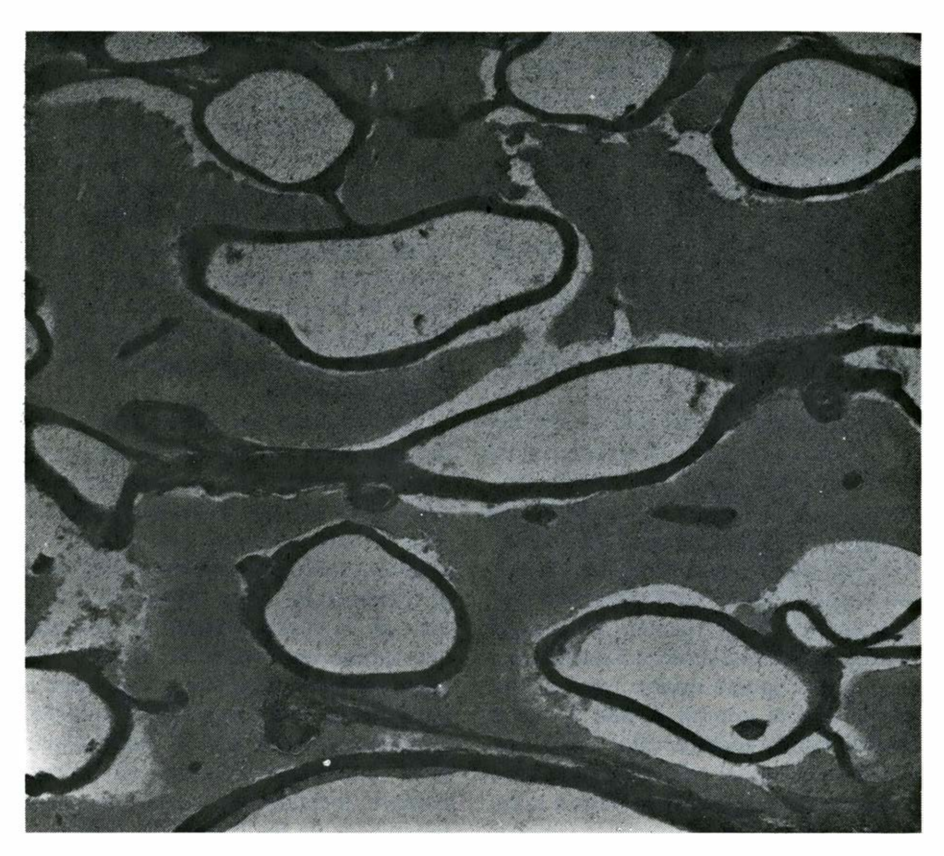
En ambas disposiciones, las arterias están bañadas por sangre venosa, aunque la sangre arterial y venosa no llegan a mezclarse. Esto no sorprende en el caso de las arterias carótidas internas, que suelen poseer paredes espesas. Pero las paredes de las arterias de la red son extraordinariamente delgadas, como pude determinar en estudios llevados a cabo en colaboración con Wendelin J. Paule y Sol Bernick, en la Facultad de Medicina de la Universidad de California meridional. ¿Por qué es tan nimia la barrera entre la sangre arterial y la sangre venosa? La respuesta a esta pregunta nos va ayudar a explicar la función principal de la *rete* carótida.

Al ser la aportación sanguínea al cerebro de gran importancia fisiológica y clínica, se comprende que la anatomía y fisiología del sistema se hayan estudiado de un modo minucioso. Los procesos metabólicos del cerebro transcurren a una gran velocidad; el órgano debe recibir, pues, un suministro de sangre con un flujo suficientemente elevado para abastecerlo de oxígeno y nutrientes y expulsar los productos de excreción. Si se corta el suministro, las células del cerebro pueden sufrir daños irreversibles; una interrupción puede ser perjudicial en cuestión de minutos o en no más de unos pocos segundos, según las circunstancias. No debe maravillarnos, pues, si la *rete* carótida, asociada íntimamente con el suministro de sangre al cerebro, ha intrigado a los investigadores durante siglos.



RETE CAROTIDA, situada en una cavidad denominada seno cavernoso, en la base del cerebro de la oveja, expuesta en sección. La sangre venosa, que sube desde la nariz y la boca, penetra en el seno por la izquierda, baña las arteriolas de la *rete* y desaparece por las venas que penetran en la vena yugular interna, a la derecha. Al mismo tiempo, desde el corazón la sangre arterial entra en la red por las ramas de la arteria carótida externa (*derecha*) y sigue por el círculo de Willis (*arriba*) después de perder parte de su calor al entrar en contacto con la sangre venosa, que se encuentra más fría.

La pregunta de si la *rete* desempeña o no un cierto papel en la regulación de la tasa de flujo sanguíneo cerebral permanece todavía sin respuesta. Lo que ahora sí se sabe es su misión en la función sensiblemente diferente de intercambiadora de calor para el enfriamiento de la sangre arterial destinada al cerebro. El trabajo que me condujo a la comprensión de esta función empezó hace unos 10 años en el laboratorio de James N. Hayward, de la Universidad de California en Los Angeles. Se sabía que la temperatura del cerebro, en el caso de vertebrados, era importante para la regulación de la temperatura corporal. A partir de ese dato nos propusimos investigar cómo se controlaba la temperatura del cerebro.



RETE CAROTIDA de un gato, vista en sección en la microfotografía. De tamaño variado, las arterias de fina pared se disponen en un lago de sangre venosa fría, afirmadas por filamentos de conectivo.

Nuestros primeros estudios arrancaron de la siguiente observación: se registraban cambios notables en la temperatura del cerebro, de hasta uno o dos grados Celsius, cuando los mamíferos pasaban de un estado de comportamiento a otro. Por ejemplo, si un conejo relajado se ponía de repente en estado de alerta, su temperatura cerebral subía, y cuando el conejo volvía a relajarse o se dormía, su temperatura cerebral descendía. Algunos investigadores habían sugerido que tales cambios en la temperatura eran causados por aumentos o disminuciones de la tasa metabólica del cerebro. Nuestros estudios demostraron que esto no era así.

Seleccionando monos y conejos como animales de experimentación, implantamos termopares muy pequeños en cada sujeto: un termopar cerca de las arterias del círculo de Willis, un segundo en el cerebro del animal y un tercero en la arteria carótida cerca de la aorta. Los hilos de los termopares estaban conectados a enchufes fijados en la cabeza de los animales; se disponían largos cables entre los enchufes y los registradores de temperatura automáticos, a fin de que pudiéramos detectar cambios en la temperatura a medida que observábamos el comportamiento del animal. Descubrimos que todos los cambios en la temperatura cerebral iban precedidos por cambios en la temperatura de la sangre arterial que abastecía al cerebro.

¿Qué originaba las modificaciones en la temperatura de la sangre arterial? Hallamos que obedecían a cambios en la tasa de pérdida de calor procedente de la piel del animal. Cuando el flujo de sangre a través de la piel aumenta, también lo hace la pérdida de calor; si el flujo disminuye, la pérdida de calor se reduce.

La dilatación y constricción de los vasos sanguíneos de la piel está controlada por el sistema nervioso autónomo, sistema que está afectado por la emoción y la actividad. Por ejemplo, cuando se excita un animal, los nervios autónomos provocan la constricción de los vasos sanguíneos de la piel; la pérdida de calor se reduce y la temperatura corporal central, incluida la temperatura de la sangre arterial, aumenta. A continuación lo hace la temperatura del cerebro. Recíprocamente, cuando un animal se relaja, el sistema nervioso autónomo permite que los vasos sanguíneos de la piel se dilaten, con lo que se incrementa la pérdida de calor a través de la piel y la temperatura central del cuerpo desciende. En el caso del conejo, la oreja constituye una de las áreas más importantes de intercambio de calor; el tegumento es delgado allí, y los tejidos están copiosamente abastecidos de sangre. En el mono, las zonas importantes de intercambio de calor son la mano y el pie; en el hombre, la región correspondiente es la mano. Ciertamente, la misma reacción autónoma se produce en los seres humanos. Cuando nos excitamos o andamos preocupados, nuestras manos tienden a enfriarse. Cuando estamos relajados, éstas se calientan.

Mientras registrábamos la temperatura cerebral y sanguínea de conejos y monos, descubrimos que la temperatura de la sangre arterial en la arteria carótida y la de la sangre arterial en el círculo de Willis eran las mismas. Surgió entonces la cuestión de si podría observarse el mismo fenómeno en otras especies. Nuestros siguientes animales de experimentación fueron gatos y ovejas, especies que a diferencia de conejos y monos sí tienen *rete* carótida. Implantamos tres termopares en los nuevos sujetos. Cuando cualquier animal se encontraba en estado relajado, hallamos que, como en el conejo y el mono, la temperatura del cerebro era mayor que la temperatura de la sangre arterial en el círculo de Willis. Si estaban activos, las temperaturas subían, como ocurría también en conejos y monos. Encontramos, sin embargo, que durante la actividad la sangre arterial del círculo de Willis permanecía más fría que la sangre en el tercer termopar, que estaba situado cerca de la aorta. De alguna manera, en gato y oveja la sangre arterial se enfriaba a medida que fluía desde el corazón al cerebro. Y eso se alejaba bastante de cuanto habíamos observado en conejos y en monos.

La única gran diferencia entre las dos series de animales de experimentación, en lo que respecta al aporte sanguíneo al cerebro, era la ausencia de *rete* carótida en monos y conejos y su presencia en gatos y ovejas. Este hecho sugirió que en nuestros nuevos sujetos la red funcionaba como un intercambiador de calor por contracorriente. Para comprobar esta hipótesis implantamos un termopar

adicional en el seno cavernoso del animal, donde la sangre venosa rodea las arterias de la *rete.* Vimos que la sangre venosa también sufría cambios de temperatura: era más fría cuando los animales estaban relajados o durmiendo y más caliente cuando estaban en alerta o excitados. Pero si abstraíamos del estado del animal, la sangre venosa estaba más fría que la sangre arterial que penetra en la red, lo que señalaba la presencia de un intercambio de calor por contracorriente.

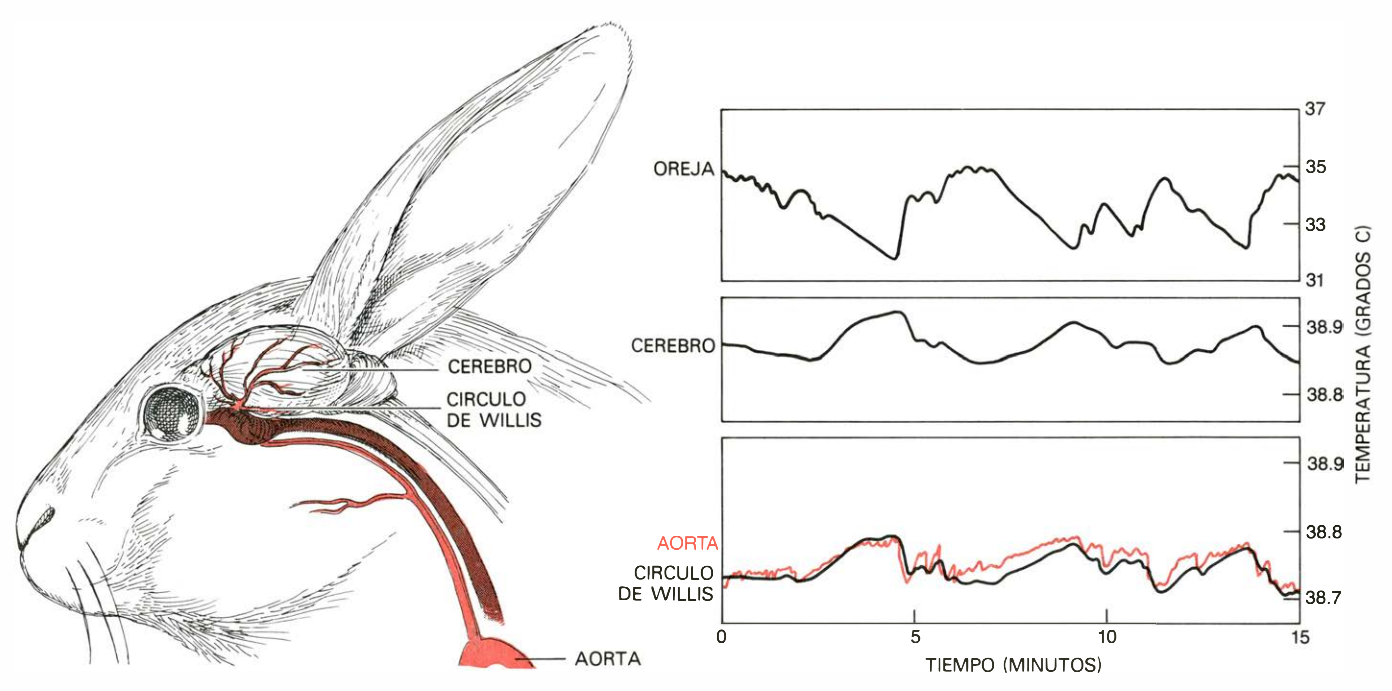
Esto nos llevó a dos nuevas cuestiones ¿Cuál era la fuente de sangre venosa en el seno cavernoso? ¿Por qué la temperatura de la sangre venosa cambiaba cuando lo hacía la actividad del animal? Respecto al primer interrogante, la sangre venosa llega al seno cavernoso procedente de varias fuentes. Una parte llega de la base del cerebro; confiábamos en que esta sangre estaría más caliente que la sangre arterial debido a que el cerebro habría añadido calor a esta sangre. Pero el caudal mayor de sangre venosa proviene de áreas externas a la cavidad craneal. Los anatomistas conocen ya, desde hace tiempo, que la sangre venosa que irriga la nariz y partes de la boca puede fluir hacia los senos intracraneales.

James H. Magilton y Curran S. Swift, de la Universidad del estado de Iowa, han estudiado la circulación venosa de la nariz con perros como sujetos de experimentación. Descubrieron que los cambios registrados en la temperatura interior de la nariz alteraban la temperatura cerebral al modificar la temperatura de la sangre venosa nasal que fluye hacia el cráneo. En un estudio anatómico realizado por nosotros obtuvimos un modelo de flujo similar en la oveja. Trabajando con cabezas de oveja, inyectamos un látex coloreado en las venas de la nariz; observamos que el látex llenaba el seno cavernoso en la base del cerebro donde está situada la *rete* carótida del animal.

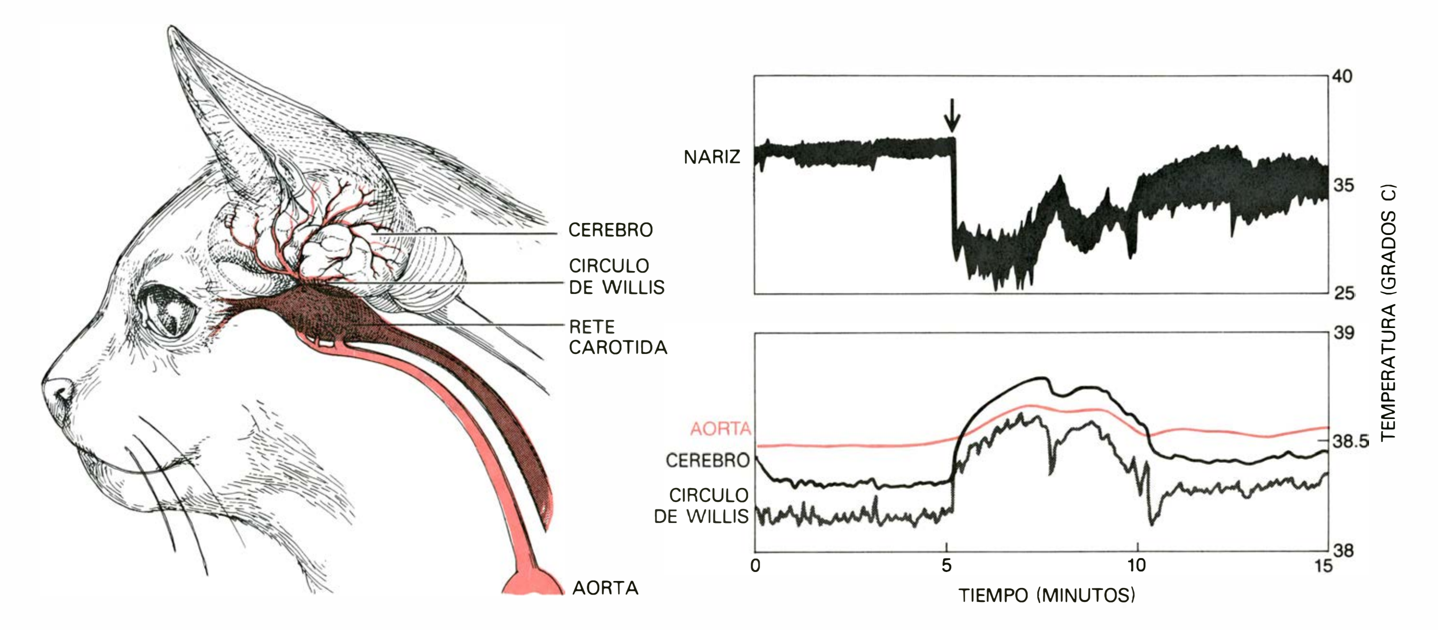
Tales estudios nos permitieron contestar a la primera cuestión: la sangre venosa que afecta a la temperatura de la red arterial procede de los conductos nasales de los animales. En cuanto a la pregunta sobre los cambios ocurridos en la temperatura, éstos reflejan la condición de la mucosa nasal: la humedad que tapiza los conductos nasales. Los vasos sanguíneos de la mucosa nasal, al igual que los vasos sanguíneos de la piel, están controlados por el sistema nervioso autónomo. Cuando un animal, hombre incluido, se relaja, los vasos sanguíneos de los conductos nasales se dilatan y la sangre venosa se enfría por evaporación de la humedad de la mucosa. Al excitarse, los vasos sanguíneos de la mucosa nasal se constriñen por los nervios autónomos. Esta acción rebaja el caudal de sangre que fluye a través de los conductos nasales y también la cantidad de sangre venosa que circula desde la nariz hasta el seno cavernoso, en el cerebro. El descenso en el suministro de sangre venosa fría disminuye el intercambio de calor entre la sangre venosa y la sangre arterial que pasa a través de la red carótida. Esto conduce a un aumento en la temperatura de la sangre arterial que llega al cerebro hasta que la temperatura de esta sangre se aproxima a la de la sangre de la región central del cuerpo.

Pudimos corroborar la validez de esta secuencia de acontecimientos con una serie de experimentos en los que se midió la temperatura de la mucosa nasal con termopares implantados en los ollares de gatos y ovejas. Cuando crecía el flujo de sangre a través de la mucosa nasal, subía la temperatura de la mucosa; cuando el flujo sanguíneo descendía, también remitía la temperatura de la mucosa. Ello estaba en perfecta consonancia con los cambios en la temperatura de la piel que habíamos observado asociados a la constricción y la dilatación de los vasos sanguíneos de la piel. (Debe advertirse que la temperatura del aire ambiental y la frecuencia respiratoria de los animales permaneció constante durante esas mediciones.)

¿Cuál es el mecanismo exacto por el que la evaporación enfría la mucosa nasal? La mucosa está siempre húmeda. En cada inspiración el animal arrastra aire relativamente seco a través de la superficie húmeda, la humedad se evapora y la mucosa enfriada enfría a su vez la sangre que circula a través de ella. Cabía esperar, pues, que un aumento en la tasa de evaporación disminuiría a continua-



CAMBIOS en la temperatura de la sangre del conejo. Tales modificaciones manifiestan estar correlacionadas con los cambios en la pérdida de calor de la oreja. La dilatación y constricción de los vasos sanguíneos de la oreja, la mayor superficie de intercambio de calor, se halla controlada por el sistema nervioso autónomo. Cuando los vasos sanguíneos de la oreja no están constreñidos, y la temperatura del aire es de 25 grados Celsius, la temperatura de la oreja se sitúa generalmente por encima de los 34 grados. Cuando los nervios autónomos constriñen los vasos sanguíneos de la oreja, la temperatura de la oreja desciende. A esa caída sigue un aumento en la temperatura de la sangre arterial que fluye desde el corazón a través de la aorta hasta el círculo de Willis. La temperatura del cerebro, algo más alta que la de la sangre en el círculo de Willis, aumenta y disminuye de un modo similar.



CAMBIOS SIMILARES DE TEMPERATURA se muestran en un gato dormido, que fue despertado una vez durante un período de 15 minutos. La mayor superficie de intercambio de calor, la mucosa de la nariz y de la boca del gato, está de igual manera bajo el control del sistema autónomo. Cuando el gato duerme, la pérdida de calor a través de la nariz es elevada y la evaporación enfría la sangre venosa que afluye al seno cavernoso. Allí, la sangre arterial más caliente que circula desde la aorta al círculo de Willis es enfriada por la sangre venosa a medida que fluye a través de la *rete* carótida. Cuando el gato se despierta (*flecha*), los nervios autónomos constriñen la circulación de la mucosa, reduciéndose la pérdida de calor nasal, así como el calor que se intercambia a través de la *rete*. La temperatura de la sangre arterial del círculo de Willis aumenta hasta casi igualar la temperatura de la sangre arterial de la aorta, en tanto que la temperatura del cerebro, normalmente más baja que la de la sangre aórtica, casi la supera. Cuando el gato vuelve a dormir, la dilatación de los vasos de la mucosa facilitan el intercambio de calor en la *rete*, y las temperaturas descienden.

ción la temperatura de la sangre venosa que fluye desde la nariz y produciría el consecuente descenso en la temperatura de la sangre arterial de la *rete* carótida.

Y eso era puntualmente lo que pasaba cuando acometimos un experimento posterior. Anestesiamos gatos y ovejas; bombeamos aire en sus narices. La sangre arterial que irrigaba la cabeza procedente del corazón permaneció a temperatura constante, aunque el aumento en la evaporación hizo que tanto la temperatura de la red como la del corazón descendieran. Cuando repetimos el experimento con un conejo como sujeto, el enfriamiento detectado no fue tan rápido. Antes bien, la primera caída de temperatura tuvo lugar en la arteria carótida, a la que siguió un descenso en la temperatura del cerebro. La secuencia indicaba que, en ausencia de *rete* carótida, la sangre venosa enfriada de la nariz debía volver al corazón y allí enfriar la sangre arterial que circula desde el corazón al cerebro antes de que la temperatura de éste quedara afectada.

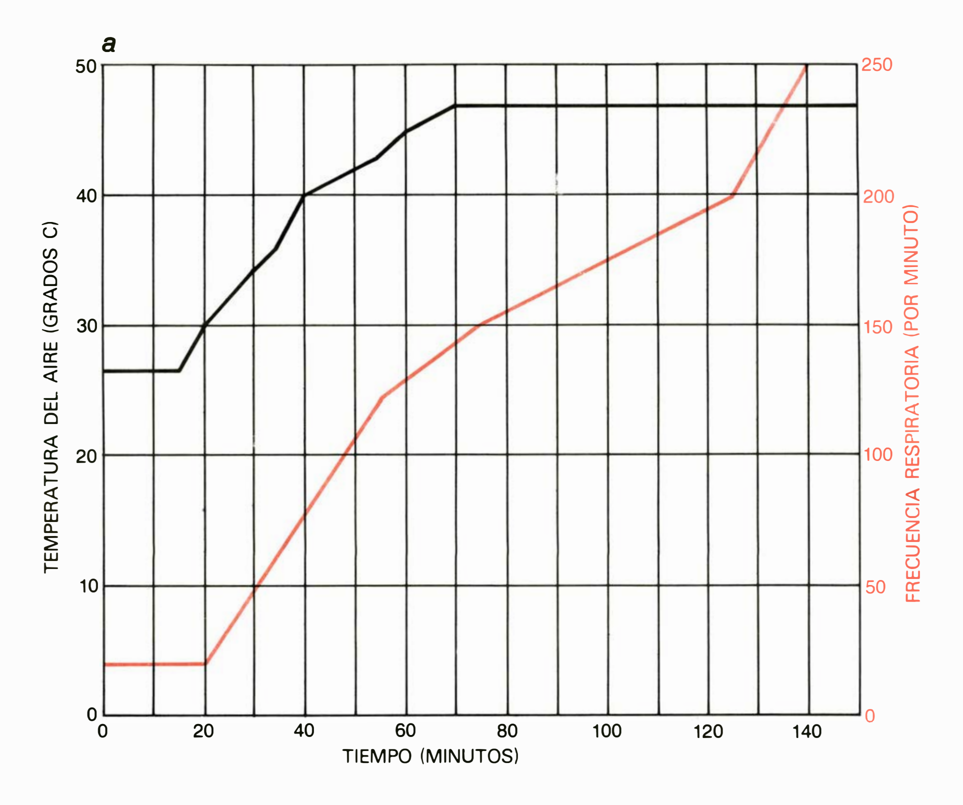
¿Tendría un incremento en la evaporación respiratoria el mismo efecto en un animal en circunstancias normales que en un animal anestesiado? Para salir de dudas introdujimos gatos y ovejas en una habitación caliente; el flujo de aire a través de sus conductos nasales tendría naturalmente que incrementarse cuando iniciaran el jadeo en respuesta al calentamiento excesivo. Cuando los animales jadeaban con una frecuencia de 250 a 300 respiraciones por minuto, la tasa de enfriamiento de la sangre arterial que circulaba a través de la red también se había incrementado. Tanto la sangre arterial de la carótida que llegaba a la red como la sangre arterial cerebral que salía de la red mostraron un aumento en la temperatura, aunque el aumento fue menor para la sangre que había pasado a través de la red.

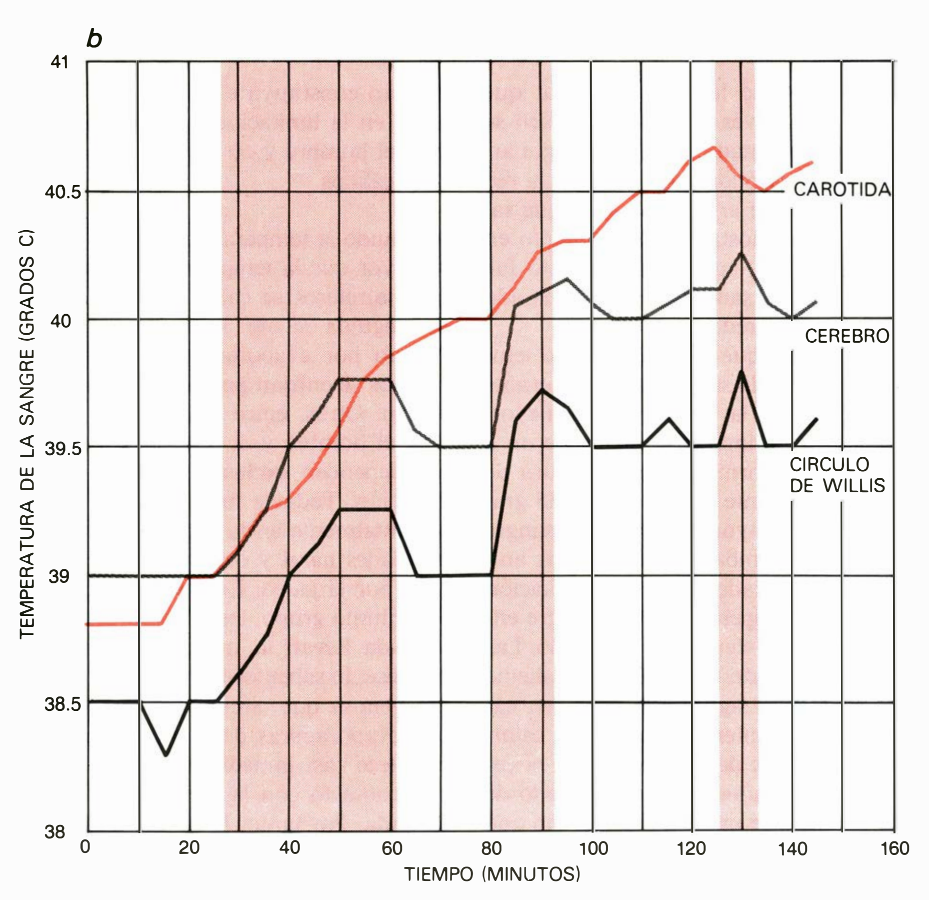
Resultó así que la diferencia de temperatura entre la sangre arterial entrante y saliente de la red se hizo mayor. Cuando los animales se mantuvieron a temperatura normal, la temperatura de la sangre entrante fue de unos 0,25 grados Celsius mayor que la de la sangre saliente. Sin embargo, cuando los animales estaban jadeando en la habitación caliente, la temperatura de la sangre entrante se había elevado en un grado. Las observaciones de la habitación caliente demostraron el significado de la *rete* carótida como intercambiador de calor. En condiciones de stress térmico, el cerebro de un animal en jadeo, dotado de *rete* carótida, se mantendrá más frío que el resto de su cuerpo.

El cerebro es especialmente sensible a temperaturas anormales. Basta una subida de cuatro o cinco grados Celsius por encima de su temperatura normal para que comiencen a dañarse las funciones cerebrales. Así, las fiebres altas en niños van acompañadas a veces de convulsiones; éstas son manifestaciones del funcionamiento anormal de las células nerviosas de un cerebro sobrecalentado. Cabe pues que la temperatura del cerebro constituya el factor más importante en la limitación de la supervivencia del hombre y otros animales en medios cálidos.

Cuando la temperatura del aire es mayor que la temperatura del cuerpo, los mamíferos se calentarían en exceso en seguida si no hubiera un enfriamiento por evaporación. Algunos mamíferos se enfrían por jadeo y por sudoración. Otros, entre los que se encuentran el hombre y el resto de los primates, dependen exclusivamente de la sudoración. Todavía hay otros que se basan totalmente en la evaporación de las cavidades nasal y oral, que se ve reforzada por el jadeo. En los mamíferos de este último grupo, los portadores de *rete* carótida llevan la mayor ventaja. Ello obedece, lo sabemos ahora, a que la sangre venosa que circula por las superficies evaporadoras de la nariz y la boca establece casi inmediatamente un estrecho contacto con las arterias de la *rete* carótida. Por tanto, la sangre que irriga al cerebro se aprovecha enteramente del enfriamiento por evaporación. Al mismo tiempo, cuando esta sangre venosa vuelve al corazón para mezclarse con la sangre venosa del resto del cuerpo del animal, aporta un enfriamiento adicional al sistema circulatorio general.

En la mayoría de los animales que jadean, la nariz y la boca están anatómicamente bien adaptadas para un eficaz enfriamiento por evaporación. Las su-



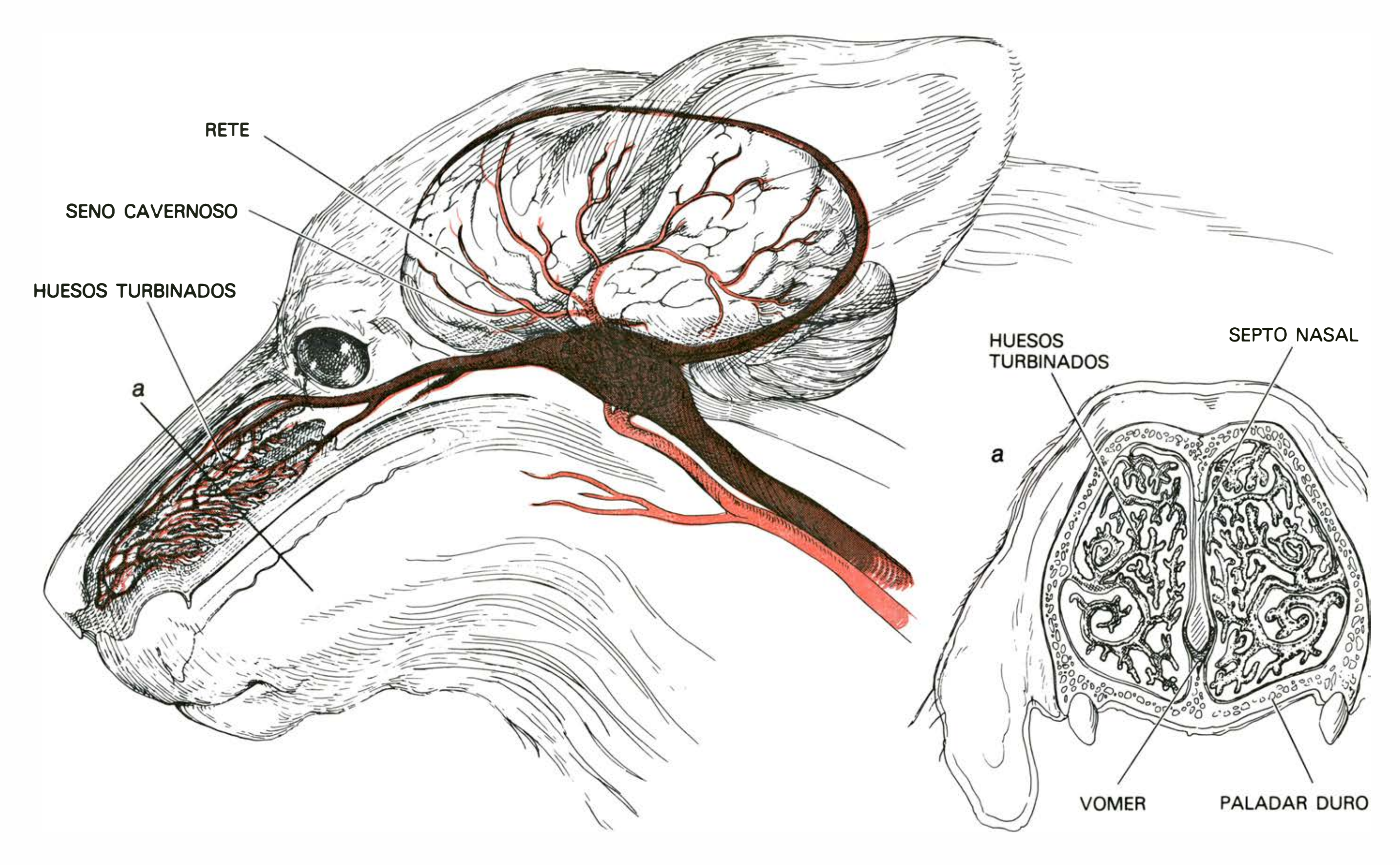


RESPUESTA AL AMBIENTE CALUROSO por parte de un mamífero protegido por la *rete* carótida. Una oveja se colocó libre en una cámara de experimentación con una temperatura ambiente de 26 grados Celsius. Durante el período de una hora, se fue elevando la temperatura del medio hasta 47 grados Celsius; se mantuvo esta temperatura por más de una hora (*a*); la frecuencia respiratoria de la oveja, de sólo 20 respiraciones por minuto en un comienzo, aumentó rápidamente y continuó incrementándose después de que la temperatura del aire se mantuviera estable. La temperatura sanguínea de la arteria carótida de la oveja subió de un modo invariable (*b*), aunque el jadeo del animal enfrió la sangre venosa circundante a la *rete* lo suficiente como para mantener la temperatura sanguínea del círculo de Willis por debajo de la sangre de la arteria carótida. Y ocurrió que el cerebro de la oveja, más caliente que la sangre de la carótida al principio, permaneció a la misma o a una temperatura más baja que la de la carótida durante la mayor parte del experimento. Cuando la oveja se levantó (*bandas coloreadas*) de su posición tumbada, la sangre del cerebro y la del círculo de Willis se volvieron más calientes.

perficies son grandes y húmedas, y las numerosas conexiones entre las arterias y las venas permiten una tasa elevada de flujo sanguíneo. Considerando primero la nariz, el complejo repliegue de huesos turbinados presenta una superficie muy grande; en algunos mamíferos, en los perros por ejemplo, se ha calculado que la superficie total de los repliegues nasales es mayor que la del resto de la superficie corporal. El interior nasal no sólo está cubierto con mucosa ricamente irrigada con venas y arterias interconectadas, sino que también alberga un órgano único en los mamíferos que jadean: la glándula nasal lateral. La glándula segrega un líquido en la mucosa, proporcionando una fuente constante de agua para la evaporación. En los perros, la tasa de secreción de la glándula nasal lateral, según los estudios de Charles Blatt, C. R. Taylor y M. B. Habal, de la Universidad de Harvard, han demostrado que es directamente proporcional a la temperatura del aire.

En cuanto a la otra región de evaporación, la boca, el agua para el enfriamiento de la mucosa que tapiza la cavidad oral la proporcionan las glándulas salivares. Estas glándulas incrementan también su secreción cuando se eleva la temperatura ambiente. Para que el proceso evaporativo alcance su máxima eficacia, la tasa ventilatoria del animal debe incrementarse al mismo tiempo que aumentan las secreciones glandulares y la tasa de flujo sanguíneo a través de la mucosa; el acto de jadear produce el necesario incremento en la ventilación.

¿Constituye la exposición a elevadas temperaturas del aire el stress térmico más severo con el que se puede enfrentar un mamífero? Para la mayoría de los mamíferos, la respuesta es negativa. Así, en mamíferos del tamaño de los gatos, perros e incluso ovejas, Taylor ha hallado que es el ejercicio, y no el aire caliente, lo que presenta la más seria amenaza de sobrecalentamiento. . En estos mamíferos de tamaño medio, las tasas metabólicas muy elevadas, producidas durante el ejercicio intenso, causan un aumento considerable de la temperatura corporal. Con fines comparativos, considérese nuestros hallazgos de que, cuando los gatos y las ovejas están en reposo bajo condiciones de calor anormales, muestran una temperatura cerebral un grado Celsius menor que la temperatura corporal. Taylor y su colega Charles P. Lyman tomaron mediciones similares de la temperatura cerebral y corporal de la gacela Thompson (antílope africano con *rete* carótida) con los animales en veloz carrera. Observaron



NARIZ DEL PERRO, en visión lateral y frontal. Contiene un conjunto de huesos turbinados que se entrelazan con gran complejidad y sostienen la mucosa nasal; esta membrana se enfría por evaporación. La sangre venosa fría, que circula desde la mucosa hasta el seno cavernoso, enfría la *rete.*

que la temperatura cerebral descendió casi tres grados Celsius por debajo de la temperatura corporal.

Este efecto refrigerador, inesperadamente notable, nos llevó a mí y a mis colegas de la Universidad de California en Riverside a sospechar que quizá la *rete* carótida no alcanzara su máxima eficiencia como intercambiador de calor excepto durante el ejercicio. En recientes estudios que hemos llevado a cabo con perros sometidos a ejercicios, se ha confirmado la conjetura, a pesar de que la *rete* carótida de los perros no está tan bien desarrollada como la de los antílopes, ovejas o gatos.

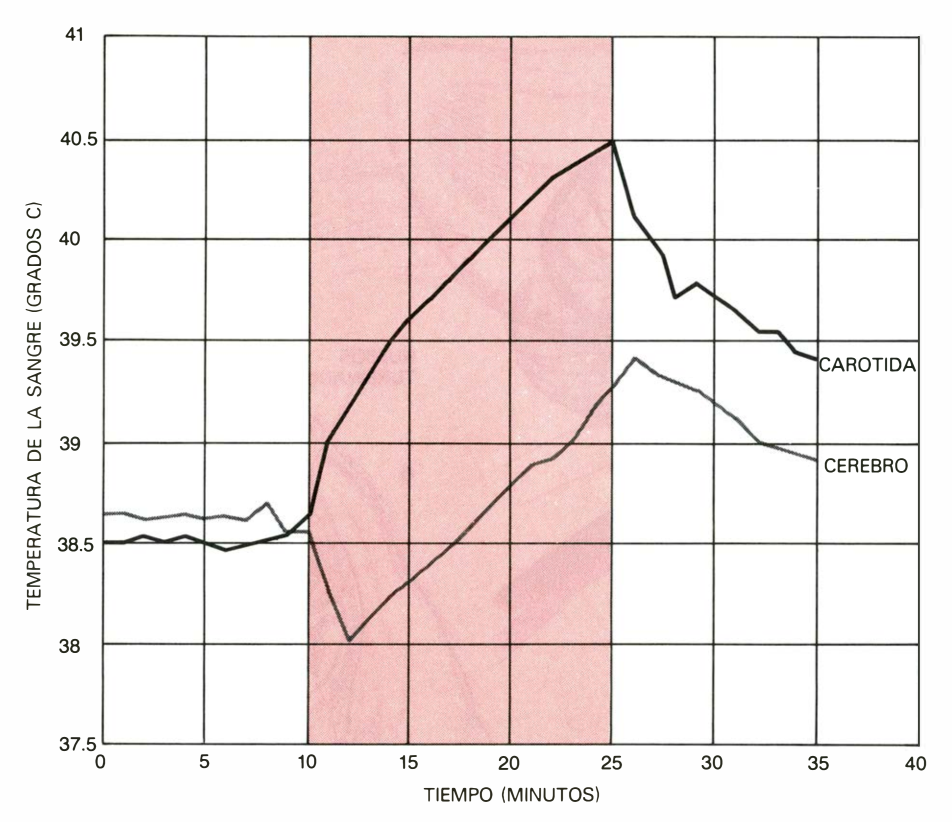
Cuando los perros fueron colocados en habitaciones calientes, apenas reaccionaron ante la temperatura anormal. Experimentaron subidas la temperatura corporal y la del cerebro, y éste permaneció levemente más frío que el cuerpo. No acabábamos de entender ese fenómeno, puesto que es sabido que los perros se cuentan entre los mamíferos más tolerantes al calor y tienen una notable capacidad para soportar largos períodos de ejercicio en condiciones desérticas extremas.

Montamos una cinta sinfín en una habitación más bien tibia, no agobiante, y medimos las temperaturas del cerebro y del cuerpo de los perros cuando corrían por la cinta sinfín. En cuanto empezaron el ejercicio aumentó rápidamente su temperatura corporal. Pero desde los primeros minutos la temperatura cerebral descendió, y permaneció unos 1,3 grados Celsius por debajo de la temperatura corporal a lo largo de todo el ejercicio; es decir: casi tres veces el enfriamiento que habíamos observado cuando los perros estaban en reposo en la habitación calurosa.

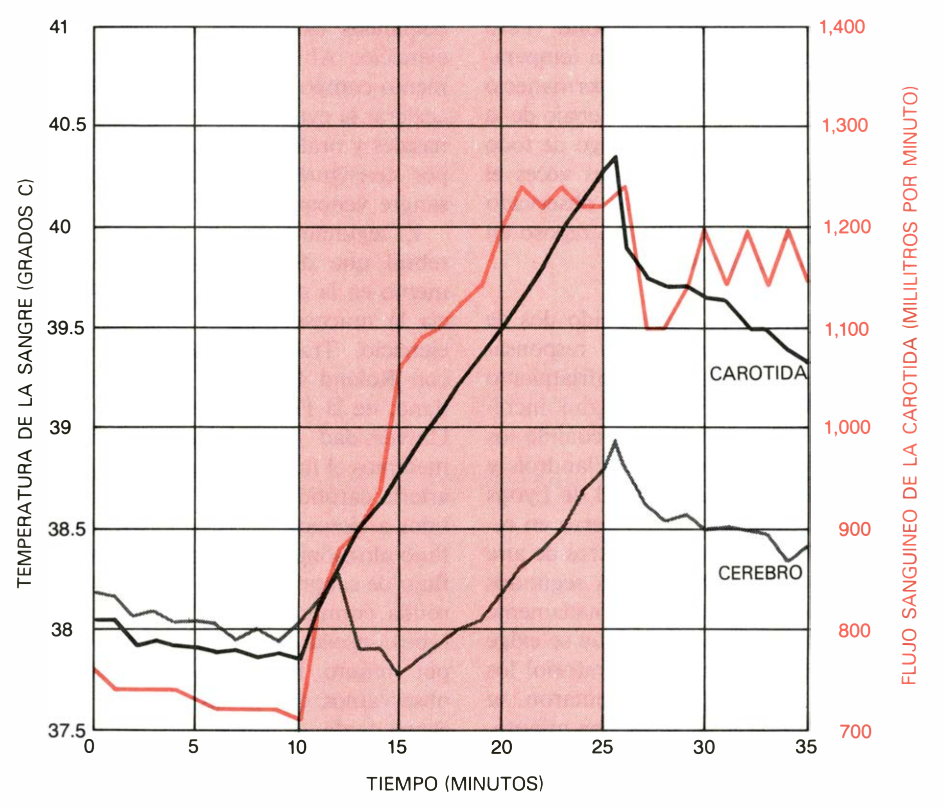
Recientemente se han medido dos de los factores fisiológicos responsables de esta elevada tasa de enfriamiento cerebral. El primero es el gran incremento en la tasa ventilatoria cuando los perros están en ejercicio. R. Flandrois y sus colegas de la Universidad de Lyons acaban de descubrir que el perro en reposo respira de cinco a seis litros de aire por minuto. En los primeros segundos de ejercicio intenso (aproximadamente el mismo nivel de ejercicio que se exige a los perros en nuestro laboratorio) los perros de Flandrois incrementaron su tasa ventilatoria a 30 litros por minuto. La tasa ascendió a continuación de un modo menos drástico, para alcanzar los 40 litros por minuto después de 15 minutos de ejercicio. Cuando el ejercicio terminó, la tasa ventilatoria descendió abruptamente. La razón fundamental del incremento en la ventilación es liberar más oxígeno y expulsar más dióxido de carbono con el objeto de satisfacer las demandas fisiológicas asociadas con el ejercicio. Al mismo tiempo, el incremento comporta el efecto secundario de acelerar la evaporación en los conductos nasales y orales del perro, descendiendo por consiguiente la temperatura de la sangre venosa.

El segundo factor de enfriamiento cerebral que debe medirse es el incremento en la tasa de flujo sanguíneo hacia la mucosa nasal y oral durante el ejercicio. Trabajando en colaboración con Roland D. Rader y William Kirtland, de la Facultad de Medicina de la Universidad de California meridional, medimos el flujo de sangre a través de la arteria carótida común en perros sometidos a ejercicio. (Usamos una sonda de flujo ultrasónico diseñada por Rader.) El flujo de sangre a través de la arteria carótida común de un perro robusto en reposo oscila entre 600 y 700 mililitros por minuto. En nuestros experimentos observamos que la tasa iniciaba su ascenso desde un comienzo y llegaba a alcanzar un nivel en exceso de 1200 mililitros por minuto.

No hay razón alguna para creer que la tasa de flujo sanguíneo cerebral crece durante el ejercicio. Por tanto, podemos suponer que la sangre adicional que fluye a través de la arteria carótida común durante el ejercicio circula hacia



LOS PERROS SOMETIDOS A EJERCICIO mantuvieron una temperatura cerebral significativamente más baja que la temperatura corporal durante un trote de 15 minutos a 7,2 kilómetros por hora (*banda coloreada*) en una cinta sinfín con una pendiente de 13 grados. Con una temperatura superior a la del cuerpo en estado de reposo, el cerebro se enfrió rápidamente al comenzar el ejercicio y permaneció a más de un grado menos durante todo el período del ejercicio. Los dos trazos para el cerebro y la carótida muestran los valores medios registrados para dos perros con cuatro pruebas.



INCREMENTO DE LA CIRCULACION durante el ejercicio, registrado en otro perro sometido a una experiencia similar, trotando en una cinta sinfín con una pendiente de 18 grados. En reposo, la sangre fluye a través de la arteria carótida común izquierda (*color*) a una tasa inferior a los 800 mililitros por minuto. Durante el período de ejercicio de 15 minutos (*banda coloreada*) la tasa de flujo sanguíneo aumentó hasta 1,2 y 1,25 litros y se mantuvo en torno a 1,1 litros durante varios minutos después de que el ejercicio finalizara. El aumento del flujo sanguíneo en la nariz y en la boca provocó un aumento en el intercambio de calor en la *rete* carótida, como revela la temperatura baja del cerebro del animal.

otras áreas diferentes del cráneo. Si sustraemos el flujo destinado al cerebro (unos 250 mililitros por minuto), podemos estimar que el flujo destinado a áreas extracraneales es de unos 400 mililitros por minuto cuando el perro está en reposo y de un litro por minuto si está en ejercicio. Puesto que la circulación arterial pasa a través de ambas arterias carótidas comunes, el flujo combinado de dos litros por minuto representa aproximadamente un sexto del total expulsado por el corazón de un perro en ejercicio.

La mayor parte de este flujo de sangre extracraneal está destinado a la mucosa de las cavidades nasal y oral del perro, donde el aumento en la tasa ventilatoria durante el ejercicio ha acelerado el proceso de enfriamiento por evaporación y las glándulas nasal lateral y salivares del perro están segregando mayores cantidades de agua para la evaporación. Por consiguiente, la combinación de factores fisiológicos, al enfriar considerablemente el cerebro de los animales que jadean y tienen *rete* carótida, proporcionan una gran tolerancia al ejercicio, en particular en ambientes calurosos.

¿Qué ocurre con los animales que jadean y no tienen *rete* carótida? Estudiamos las temperaturas cerebral y corporal de conejos domésticos cuando los animales estaban corriendo. Hallamos que su temperatura cerebral se incrementaba en paralelo con su temperatura corporal. El jadeo del animal, sin embargo, puede hacer algo para compensar el aumento en la temperatura cerebral, como han demostrado M. Caputa y sus colegas de la Universidad Nicholas Copernicus de Torun, Polonia. Al medir la temperatura de la sangre venosa que retorna de la nariz en el conejo en jadeo, vieron que se enfriaba por evaporación. En consecuencia, un enfriamiento por conducción rebajaba la temperatura de las regiones del cerebro del conejo adyacentes a la sangre venosa fría; estas partes del cerebro demostraron estar a una temperatura 0,5 grados Celsius menor que la del cuerpo del conejo.

En la naturaleza, este factor de compensación parece ser trivial. Vaughan H. Shoemaker y sus colegas de Riverside han recurrido a la radiotelemetría para medir la temperatura corporal de "jackrabbits" (especie americana de conejo, del género *Lepus*) en condiciones naturales en el desierto de Mojave. En un día caluroso de verano, el jackrabbit en reposo tiene una temperatura corporal de unos 41 grados Celsius. Cuando el animal es expulsado de su lugar de reposo y perseguido, muestra un rápido incremento en su temperatura corporal. Entre cinco y 10 minutos de ejercicio

puede aumentar su temperatura corporal por encima de los 43 grados. Knut Schmidt-Nielsen y sus colegas de la Duke University han demostrado que, si la temperatura corporal del jackrabbit aumenta por encima de los 44 grados, el animal muere. Es inmediato, pues, que un jackrabbit que no pueda descansar y enfriarse por cortos periodos en su carrera no podrá sobrevivir por largo tiempo.

El estudio de la función intercambiadora de calor de la *rete* carótida en los animales con jadeo ha suscitado nuevas cuestiones de interés en torno a la regulación general de la temperatura en los mamíferos. Desde hace tiempo se sabe que hay una región sensible al calor en el tronco cerebral de los mamíferos y que, midiendo la temperatura de la sangre arterial que fluye desde el centro del cuerpo al cerebro, esta región actúa a buen seguro como un termostato que controla la temperatura corporal central. Esta suposición parece válida para los mamíferos que carecen de *rete* carótida; en éstos, la sangre arterial no cambia de temperatura cuando pasa del cuerpo al cerebro. Ahora bien, ¿es sensible el termostato del tronco cerebral en un mamífero dotado de *rete* carótida?

Lo que se mide es, cuando menos, ambiguo. Por ejemplo, durante el ejercicio la temperatura registrada por el termostato difiere bastante de la temperatura real del centro corporal; la temperatura del cerebro y la corporal cambian en direcciones opuestas en el inicio del ejercicio y permanecen a distintos niveles durante la mayor parte del período del ejercicio. ¿Cómo conoce entonces el termostato del cerebro cuál es la temperatura corporal? ¿Está la temperatura cerebral y la corporal reguladas independientemente en los mamíferos dotados de *rete* carótida?

Podemos conjeturar que, en tales animales, células nerviosas sensibles a la temperatura de partes del cuerpo distintas del cerebro cumplen un papel importante en la regulación de la temperatura durante el ejercicio. Por ejemplo, tales sensores existen en la medula espinal y en la cavidad abdominal y han demostrado jugar un papel termorregulador en algunos mamíferos. En mamíferos con *rete* carótida, el envío de estímulos desde estos sensores extracraneales a la región del cerebro que controla la temperatura corporal puede ser mayor durante el período de ejercicio que durante el período de reposo. La respuesta correcta a tales preguntas se sabrá a partir de estudios posteriores de termorregulación en una amplia variedad de mamíferos.

Problemas intrínsecamente difíciles

La resolución de ciertos problemas de cómputo exigiría un ordenador tan grande como el Universo, trabajando un tiempo al menos igual a la edad del Universo. No obstante, tales problemas son en principio resolubles

Larry J. Stockmeyer y Ashok K. Chandra

Ciertos problemas matemáticos (especialmente en lógica y en teoría de números) son en principio resolubles, pero por métodos que rebasan con mucho la capacidad del más potente ordenador imaginable. Ciertas propiedades intrínsecas del Universo impondrán siempre límites a la velocidad y tamaño de los ordenadores. El más potente ordenador que pueda concebirse no podrá ser mayor que el Universo conocido (cuyo diámetro es inferior a 100.000 millones de años luz), no podrá estar formado por componentes menores que un protón (diámetro 10^{-13} centímetros) y no podrá transmitir información más velozmente que la luz (3×10^8 metros por segundo). Con estas limitaciones, tal computador podría estar formado por un máximo de 10^{126} componentes elementales. Albert R. Meyer, del Instituto de Tecnología de Massachusetts, y uno de nosotros (Stockmeyer) han demostrado que este ordenador ideal, independientemente del ingenio con que sea diseñado y de la perfección de su programa, necesitaría por lo menos 20.000 millones de años para resolver ciertos problemas matemáticos que, al menos en principio, se sabe que admiten solución. Como probablemente el Universo no alcanza a tener 20.000 millones de años de antigüedad, parece razonable afirmar que tales problemas desafían el análisis por ordenador.

Problemas capaces de saturar el ordenador ideal durante al menos 20.000 millones de años son sin duda los casos más extremos de problemas no-computables ("intractable problems"). Sin llegar a casos tan extremos, existen problemas cuya resolución precisaría años de funcionamiento de los más modernos y veloces ordenadores, independientemente del método empleado en su resolución. Hay por otra parte problemas para los que no se han encontrado métodos eficientes de solución, si bien es posible que tales métodos existan. En estos últimos años, matemáticos y especialistas en computación han formalizado la noción de no-computabilidad ("intractability"), identificado con precisión diversas categorías de problemas no computables y demostrado muchos teoremas para cada una de estas categorías de problemas. El resultado de todos estos trabajos suscita cuestiones fundamentales sobre las capacidades y limitaciones de los sistemas de proceso de datos.

Según parece, suelen presentarse problemas no computables en juegos como ajedrez, damas, go y otros. Supongamos que dos personas se enfrentan en una partida de damas, y que solamente queden cuatro piezas sobre el tablero: tres damas rojas y una dama negra. Si la próxima jugada le corresponde a las piezas rojas, éstas disponen de una estrategia que antes o después les dará la victoria, por mucho que se esfuercen las negras en impedirla. Si, en lugar de cuatro piezas, quedasen todavía en el tablero alrededor de una docena, en la mayoría de los casos no sería evidente que las rojas dispusieran (o no) de una estrategia invencible. En principio, es posible hallar la respuesta a la "fuerza bruta", analizando todas las posibles continuaciones del juego por las rojas, y todas las posibles contramedidas de las negras.

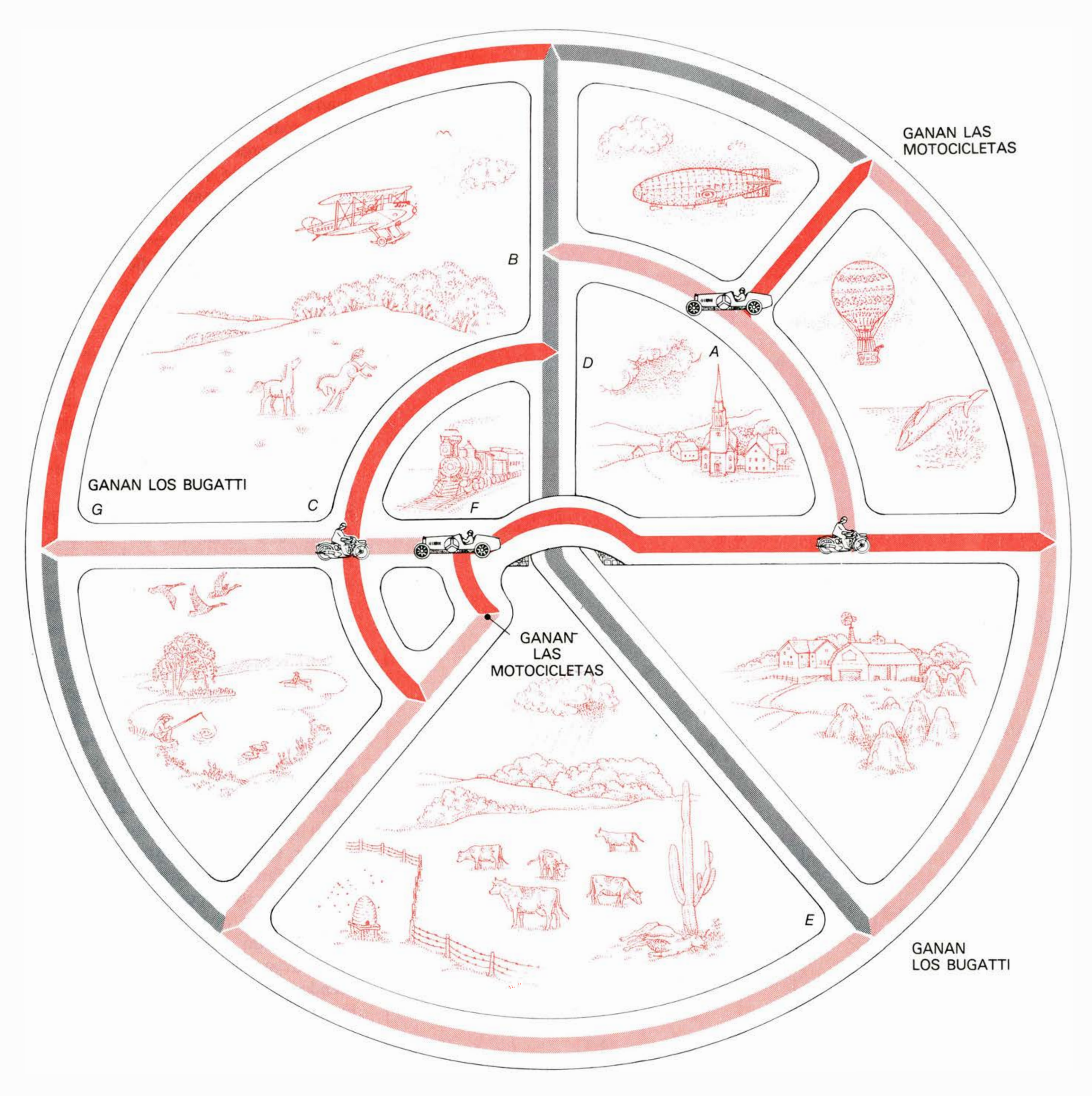
Un método que permite estudiar sistemáticamente todas las estrategias y contraestrategias posibles consiste en construir un "diagrama de juego", formado por una colección de puntos conectados mediante flechas. Cada punto representa una posición del juego, es decir, la descripción detallada del emplazamiento de todas las piezas más la información sobre quién moverá en la próxima jugada. Cada posible jugada a partir de una posición dada se representa mediante una flecha que conecta los correspondientes puntos. Cuando una posición no contiene piezas negras, se tiene una posición ganadora para las rojas, y es señalada como tal. Si la posición no contiene piezas rojas, es señalada como posición ganadora para las negras.

Los demás puntos (posiciones) del diagrama pueden ahora identificarse como posiciones ganadoras bien para las rojas, bien para las negras, sin más que basarse en unas cuantas reglas sencillas. Si el rojo puede mover desde la posición *A* hasta la *B*, y se ha establecido ya que la posición *B* da la victoria a las rojas, entonces *A* puede señalarse también como posición ganadora de las rojas. Por idéntica razón, si desde la posición *A* las rojas solamente pueden mover hasta posiciones que han sido ya identificadas como de victoria de las negras, entonces también *A* puede etiquetarse como posición ganadora de las negras. El procedimiento de rotulado prosigue, hasta que no puedan ya marcarse de una u otra forma nuevos puntos. Las posiciones no marcadas que puedan quedar corresponden a empates. Prestando atención para evitar que pueda haber más de un punto que represente una misma posición, el diagrama puede construirse y marcarse en un tiempo finito. De esta forma, toda posición inicial puede ser analizada y clasificada como de "tablas", de "victoria para rojas" o "victoria para negras".

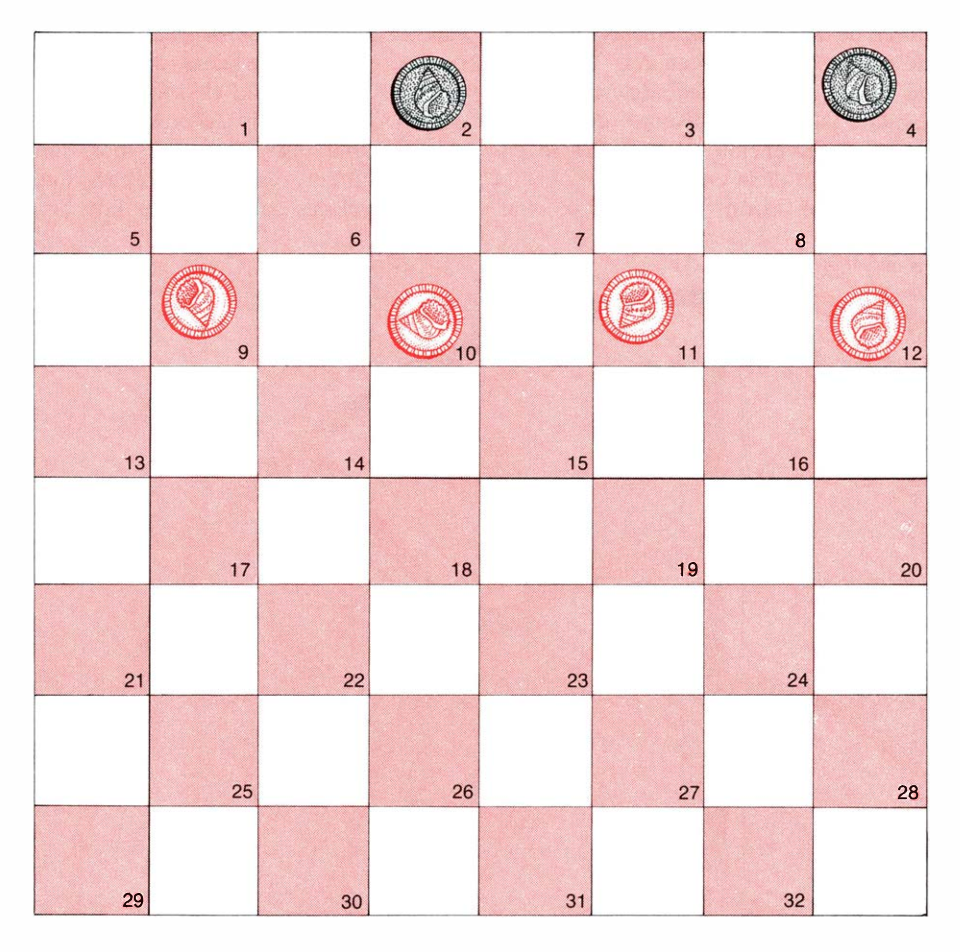
Estos métodos de análisis se llaman algoritmos: sistemas bien determinados de instrucciones precisas que pueden aplicarse mecánicamente y que permiten obtener las soluciones de casos concretos de problemas tipo. Cuando es necesario resolver un problema mediante ordenador, es indispensable preparar un algoritmo, pues la máquina solamente puede obedecer instrucciones enunciadas sin ambigüedad y en forma de algoritmo. Tenemos ejemplos de algoritmos en los procedimientos de sumar y de multiplicar números enteros que se enseñan en la escuela primaria. Aplicando mecánicamente dichos procedimientos a cualquier par de números enteros, los corres-

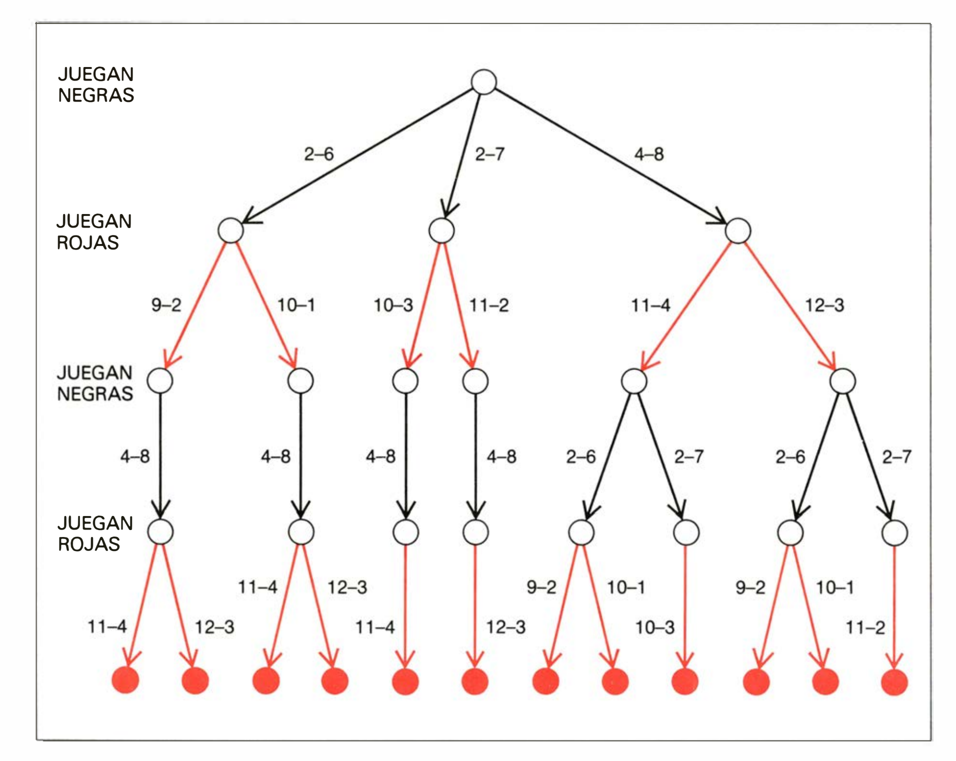
pondientes resultados son siempre correctos. Al construir un diagrama de juego se está dando un algoritmo para analizar cualquier situación de una partida de damas. Algoritmos sustancialmente idénticos permitirían analizar juegos como el ajedrez y el go, que no se sirven de dados, naipes, ruletas ni demás elementos que impliquen azar. En general, el algoritmo de juego determina qué jugador (si lo hay) dispone de alguna estrategia invencible al encontrarse en una posición dada cualquiera.

¿Por qué razón, entonces, no son capaces los ordenadores electrónicos de realizar partidas perfectas de damas o ajedrez? La respuesta es, sencillamente, que en situaciones típicas de estos juegos el diagrama está formado por tan enorme número de puntos, tantos, que los más veloces computadores hoy dis-



"ROADBLOCK" es un juego de mesa que se desarrolla sobre una red de carreteras, pintadas cada una de un color elegido entre tres. Dos jugadores, que actúan por turno, tienen cada uno una flota de coches, que inicialmente ocupan ciertas posiciones designadas. Llegada su vez, cada jugador mueve uno de sus coches hasta una nueva intersección viajando por una carretera del mismo color que se encuentre libre de vehículos. El primer jugador que consiga llevar uno de sus coches hasta un cruce estipulado de antemano gana la partida. En la ilustración, el jugador provisto con los Bugattis podría mover uno de ellos desde el cruce *A* hasta el cruce *B*, amenazando vencer en su próxima jugada. Si el contrario intentase bloquear la carretera que une *B* con *E*, llevando una de sus motocicletas desde *C* hasta *D*, el primer jugador vencería moviendo un Bugatti desde *F* hasta *G*. Esto significa que si el jugador que maneja los Bugattis es el primero en mover, entonces dispone de una estrategia invencible. En general no es fácil analizar posiciones arbitrarias del "roadblock". Existen algoritmos, esto es, rutinas aplicables mecánicamente, que permiten determinar cuál de los jugadores dispone de estrategia invencible en cualquier posición dada, pero todos ellos son "ineficientes": el tiempo necesario para ejecutar uno cualquiera de ellos crece exponencialmente, con la misma rapidez que c^n, siendo c una constante mayor que 1 y n el número de cruces de la posición a estudiar. Si bien no todo algoritmo necesitará una cantidad exponencialmente creciente de tiempo en hallar una estrategia invencible para un jugador en toda posición del juego, para cada algoritmo hay infinidad de posiciones cuya evaluación exige unos tiempos exponencialmente crecientes.





POSICIONES DE UNA PARTIDA de damas (*arriba*), analizadas mediante "un diagrama de juego" (*abajo*), cuya finalidad es determinar si la conclusión de un juego perfecto será victoria para las rojas, victoria de las negras, o empate. Cada punto del diagrama corresponde a una posición de las damas; el primer punto denota la posición ilustrada en la figura. Cada posible jugada está representada por una flecha rotulada con un par de números; el primero indica la casilla que ocupa la ficha, el segundo, la nueva casilla que podría ocupar. Como todos los puntos de la última fila (*en rojo*) corresponden a posiciones que no contienen piezas negras, la posición inical presentada es ganadora para las rojas. Para analizar posiciones cualesquiera no puede utilizarse el método de construcción de diagramas de juego, porque el tamaño del diagrama crece exponencialmente en función del número de piezas.

ponibles necesitarían años de funcionamiento para examinarlos todos. Los programas de ordenador que juegan al ajedrez o a las damas no engendran el diagrama completo, sino únicamente pequeñas partes del mismo, de modo muy semejante a como hace un jugador humano cuando "anticipa" el desarrollo del juego. Por este motivo los programas no desarrollan partidas perfectas, aunque algunos de ellos lo hacen francamente bien en comparación con jugadores humanos.

Por improbable que pueda ser, los matemáticos no han descartado completamente la posibilidad de que algún día lleguen a prepararse programas de ordenador capaces de desarrollar partidas de ajedrez perfectas, y más eficientes que los procedimientos de "fuerza bruta", de construcción del diagrama de juego completo. De hecho, se han podido preparar algoritmos sumamente perspicaces y eficientes para diferentes juegos. Uno de ellos es el "nim". En este juego se tienen tres pilas de fichas o cuentas de cualquier tipo; los jugadores van por turno alternativo retirando el número de fichas que deseen del montón que prefieran (aunque están obligados a retirar una ficha por lo menos). El jugador que consiga quedarse con la última ficha gana la partida.

Aunque las situaciones que se presentan en el juego de nim podrían analizarse satisfactoriamente construyendo su diagrama de juego, hay un algoritmo mucho más eficiente, que permite analizar cualquier situación sin más que efectuar unos cuantos cálculos sencillos, no más difíciles que una suma de números [*véase la ilustración de la página 96*]. Tan eficiente es este algoritmo que un jugador humano podría evaluar situaciones correspondientes a miles de cuentas en unos cuantos minutos, sin más ayuda que lápiz y papel. En cambio, para evaluar tales situaciones construyendo el diagrama de juego se necesitarían varios años aún disponiendo del más veloz computador. El juego de nim permite así poner de manifiesto una importante cuestión, a saber: que para resolver un mismo problema puede haber lo mismo algoritmos rápidos que algoritmos lentos.

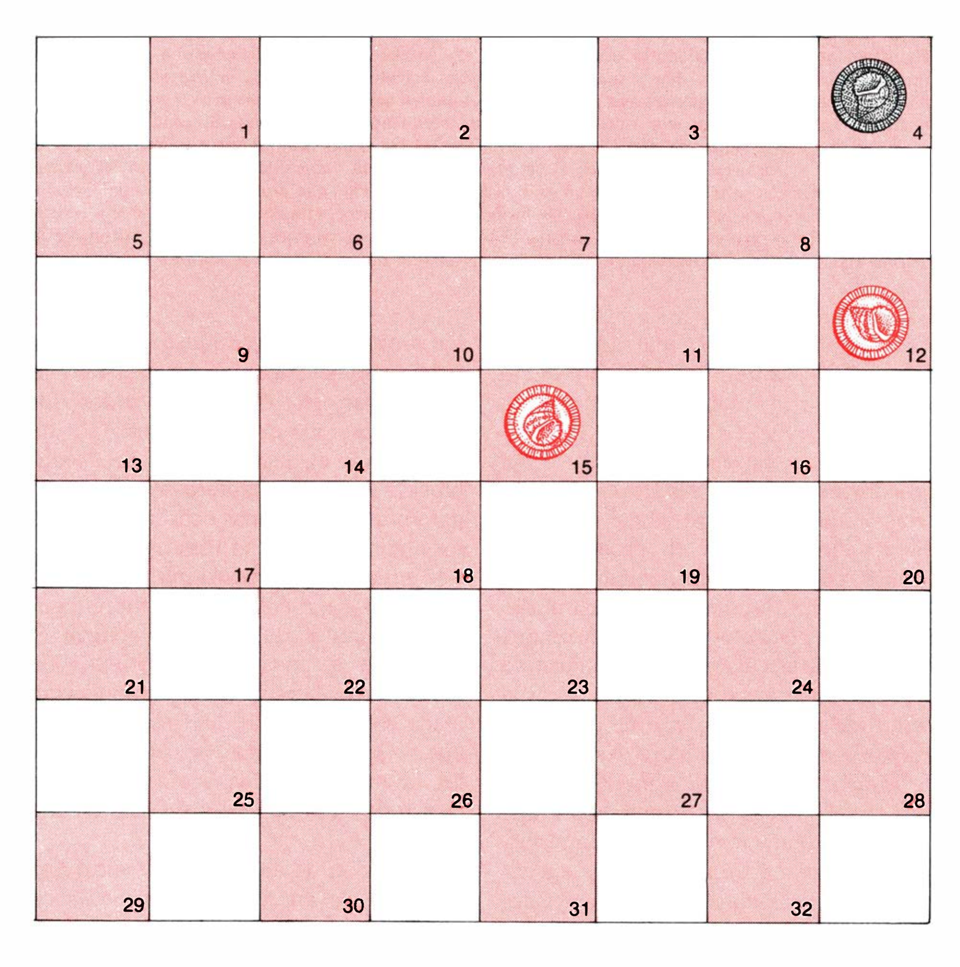
La eficiencia de un algoritmo se mide estudiando la forma en que va creciendo el tiempo requerido para ejecutarlo conforme aumenta el tamaño de la entrada. La entrada es el caso concreto del problema tipo supuestamente resuelto por el algoritmo; por ejemplo, una posición concreta de una partida de

damas o de nim. Que las damas no podrán analizarse eficientemente construyendo un diagrama de juego se demuestra observando que el número de puntos del diagrama crece exponencialmente en función del número de piezas que componen la posición a estudiar. Dicho de otra forma, el diagrama correspondiente a una posición con n piezas consta de al menos c^n puntos, siendo c una constante mayor que 1. Pequeños aumentos en el número de damas provocan aumentos explosivos del número de puntos del diagrama, y en consecuencia, del tiempo requerido para construir y marcar el diagrama completo. Aunque en el juego de damas ordinario n nunca sea mayor que 24, el algoritmo podría utilizarse para versiones arbitrariamente grandes del juego, por ejemplo, en tableros de 20 por 20, con 60 piezas. Evidentemente, el crecimiento exponencial del tiempo de ejecución de este algoritmo de "fuerza bruta" sigue cumpliéndose en estos nuevos y mayores valores de n. En contraste, el tiempo de ejecución del algoritmo rápido de análisis del nim crece en función lineal de n (funciones como $6n$ o $10n$), siendo n el número de cuentas de la posición. El rápido crecimiento de las funciones exponenciales –frente al relativamente lento de las funciones lineales– es la causa de la dispar eficiencia de ambos algoritmos.

Es opinión generalizada que los algoritmos cuyos tiempos de ejecución son funciones exponenciales, es decir, crecen al ritmo de c^n, no tienen interés práctico. Cuando c sea solamente un poco mayor que 1 –pongamos por caso, 1,1– tales algoritmos pueden ser útiles mientras la entrada sea muy pequeña. Aun así, el crecimiento exponencial sigue siendo explosivo para entradas mayores; el valor concreto de c solamente influye en el momento en que el crecimiento exponencial manifiesta tan claramente sus efectos que el algoritmo pierde toda utilidad. Si se desea resolver mediante ordenador un problema para el que solamente se conocen algoritmos exponenciales, hay buenos motivos para buscar algoritmos más eficientes. Si se alcanzase a descubrir un algoritmo mucho más rápido, los recursos invertidos en descubrirlo pagarán elevados intereses cada vez que sea necesario utilizar el algoritmo.

No obstante, siempre existe el riesgo de que tales inversiones no entreguen dividendos, porque puede ocurrir que no existan para ese problema algoritmos eficientes. Es posible que por su complejidad, el problema sea "exponencial por naturaleza": esto es, que todas sus soluciones algorítmicas exijan tiempos crecientes exponencialmente. Tal posibilidad ha motivado que los matemáticos hayan intentado probar que ciertos problemas son inherentemente exponenciales. Para muchos problemas, como en el análisis de situaciones de juego en ajedrez o damas, los algoritmos más rápidos que se conocen tienen tiempos de crecimiento exponencial, si bien no se ha demostrado todavía que sea imposible la existencia de algoritmos eficientes para ellos.

La primera demostración del carácter inherentemente exponencial de un problema fue propuesta por Meyer y uno de los autores (Stockmeyer) en 1972, con respecto a un problema de una rama de las ciencias de computación conocida como teoría de lenguajes formales. Pronto se dieron demostraciones parecidas para problemas de otras disciplinas matemáticas. En 1976 los autores demostraron que los análisis algorítmicos de diversos juegos son inherentemente exponenciales. Uno de estos juegos es el llamado "roadblock", que se desarrolla sobre un tablero en el cual se ha dibujado una red de carreteras que se cortan. Cada carretera está pintada de un color, y hay tres colores. Cada uno de los dos jugadores, que actúan por turnos alternativos, dispone de una colección de marcadores, inicialmente situados en ciertas intersecciones designadas. Al llegar su vez, el jugador va recorriendo las carreteras con uno de sus marcadores, hasta otra intersección, debiendo cumplir el requisito de que todas las carreteras recorridas sean de un mismo color y



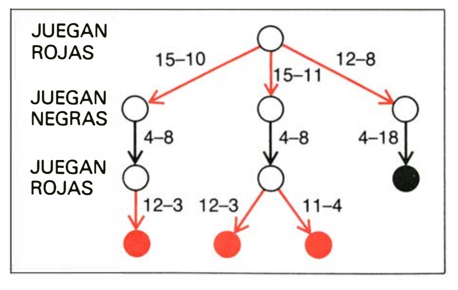
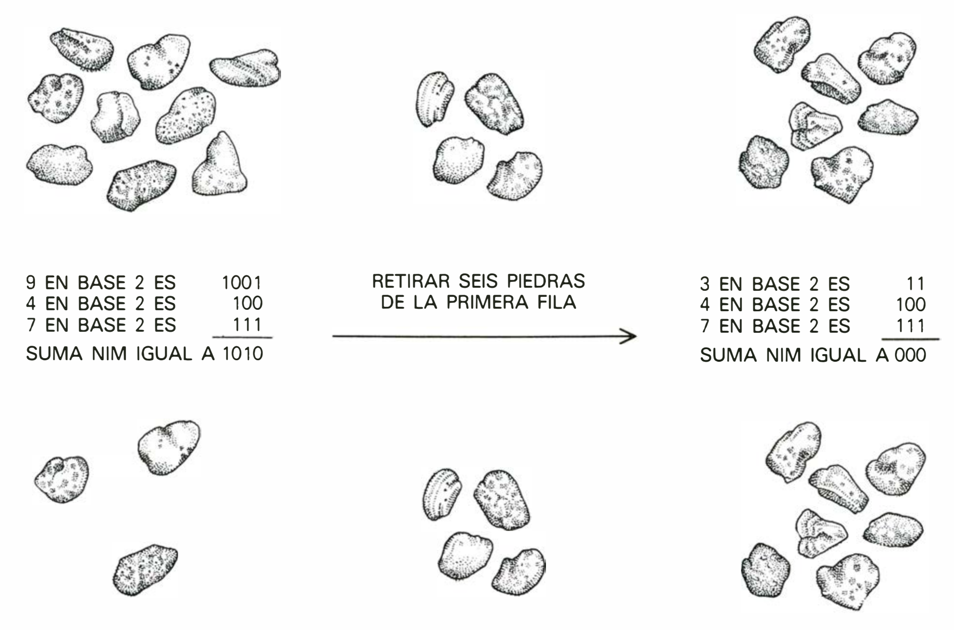


DIAGRAMA DE JUEGO (*abajo*) **de una posición de una partida de damas** (*arriba*)**; tanto puede tener posiciones finales con victoria de las negras** (*puntos negros*)**, que no contienen piezas rojas, como posiciones finales con victoria roja** (*puntos rojos*)**, que no contienen piezas negras. Si en la posición indicada le correspondiese jugar a las rojas, éstas podrían asegurarse la victoria moviendo desde el cuadro 15 bien al cuadro 10, bien al cuadro 11. Por tanto, con la posición inicial dada, las rojas tienen una estrategia invencible.**



EL NIM es un juego bipersonal donde se usan cuentas o fichas apiladas en montones. Llegado su turno, cada jugador retira de una de las pilas el número de fichas que desee. (Ha de retirar obligatoriamente una ficha al menos.) El jugador que se lleva la última cuenta gana la partida. Si bien la partida puede analizarse exhaustivamente mediante un diagrama de juego, el algoritmo es ineficiente. Hay otro algoritmo que permite analizarlo muchísimo más eficientemente. Para determinar si el jugador de turno en la posición de la figura superior tiene alguna estrategia vencedora, se expresa en notación binaria el número de cuentas de cada montón, y se anotan en columna estos números (*centro*). Seguidamente, se calcula la "suma nim" de estos números, escribiéndose un 1 debajo de cada columna que contenga un número impar de cifras 1, y un 0 debajo de las columnas con número par de cifras 1. Si la "suma nim" es cero, es posible vencer al primer jugador. Si la "suma nim" no es cero, el primer jugador puede siempre vencer, jugando en cada turno a una posición que produzca "suma nim" igual a cero. Por ejemplo, en la figura izquierda de la ilustración superior, el primer jugador debería retirar seis cuentas de la primera pila para alcanzar la posición que se muestra abajo. El algoritmo de la "suma nim" funciona eficientemente para cualquier número de cuentas que se usen en los montones.

no contengan otros marcadores. El jugador que antes llegue a un cierto cruce estipulado como "meta" gana la partida.

Dada una posición cualquiera del "roadblock" (una red arbitraria de carreteras y una disposición inicial arbitraria de los marcadores) el problema de análisis consiste en determinar cuál de los jugadores dispone de una estrategia invencible. Los autores demostraron que el tiempo de ejecución de todo algoritmo capaz de analizar el juego ha de crecer al ritmo de c^n, siendo c una constante mayor que 1, y n el número de cruces de la posición. Es necesario insistir en que esta conclusión no es simplemente una buena conjetura comprobada empíricamente ensayando unos cuantos algoritmos, y confirmada por los fracasos para encontrar un algoritmo eficiente; por el contrario, la conclusión es resultado de una demostración matemática.

También es importante darse cuenta de que la demostración no prueba que cada posición con n intersecciones obligará a todo algoritmo a consumir tiempos con crecimiento exponencial. Lo mismo que ocurría en la posición de tres damas contra una, cuyo análisis es sencillo, el "roadblock" contiene posiciones fáciles de analizar. La mayor parte de los problemas de este juego no llegan a ser lo suficientemente complejos como para dejar en ridículo a todos los algoritmos; lo que realmente se ha demostrado es que el análisis del "roadblock" sí es lo bastante complejo como para burlar en algunas ocasiones a todos los algoritmos. Dicho más exactamente: no importa qué algoritmo pueda tomarse, el juego contendrá siempre algunas posiciones difíciles que obligan al algoritmo a consumir en el análisis tiempos con crecimiento exponencial. Nada sabemos sobre la frecuencia con que se presentan estas posiciones difíciles. Lo más que podemos afirmar es que, para todo valor de n, existe al menos una de ellas.

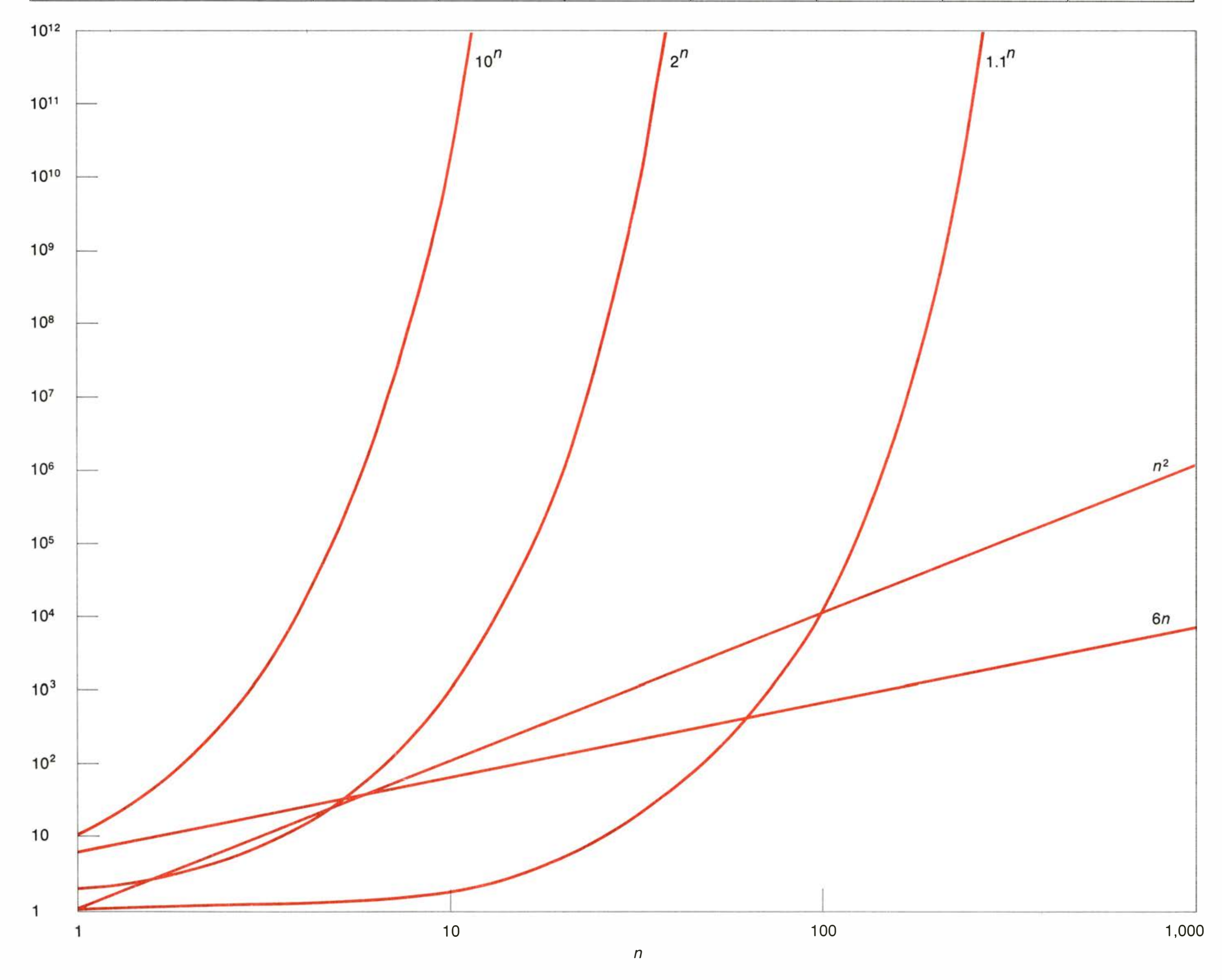
La noción de algoritmo cristalizó por primera vez en términos matemáticos precisos hacia los años 30, al intentar los lógicos de aquel tiempo esclarecer las condiciones precisas para generar mediante procesos rutinarios demostraciones de teoremas matemáticos. El brillante matemático británico A. M. Turing fue uno de los primeros en formular con precisión las características de tales procesos. Turing postuló que todo algoritmo ha de poderse ejecutar por una máquina formada por una cinta infinita de papel, dividida en cuadrículas, más un dispositivo de impresión capaz de imprimir marcas en las cuadrículas y de borrar las ya escritas, y por un sensor que determine si una cuadrícula dada está marcada o no. Conocida por el nombre de máquina de Turing, este dispositivo imaginario puede ser programado para hallar la solución de problemas, ejecutando para ello un número finito de acciones de impresión y borrado.

Lo notable de la máquina de Turing es que, a pesar de su extrema sencillez, ningún otro dispositivo de cómputo que actualmente se conozca la aventaja en la resolución de problemas. Concediéndole suficiente tiempo, la máquina de Turing puede, en principio, resolver todo problema que pueda resolver el computador más perfeccionado. En consecuencia, se ha aceptado universalmente que la condición necesaria y suficiente para que un problema pueda ser resuelto por vía algorítmica es que pueda ser resuelto por una máquina de Turing.

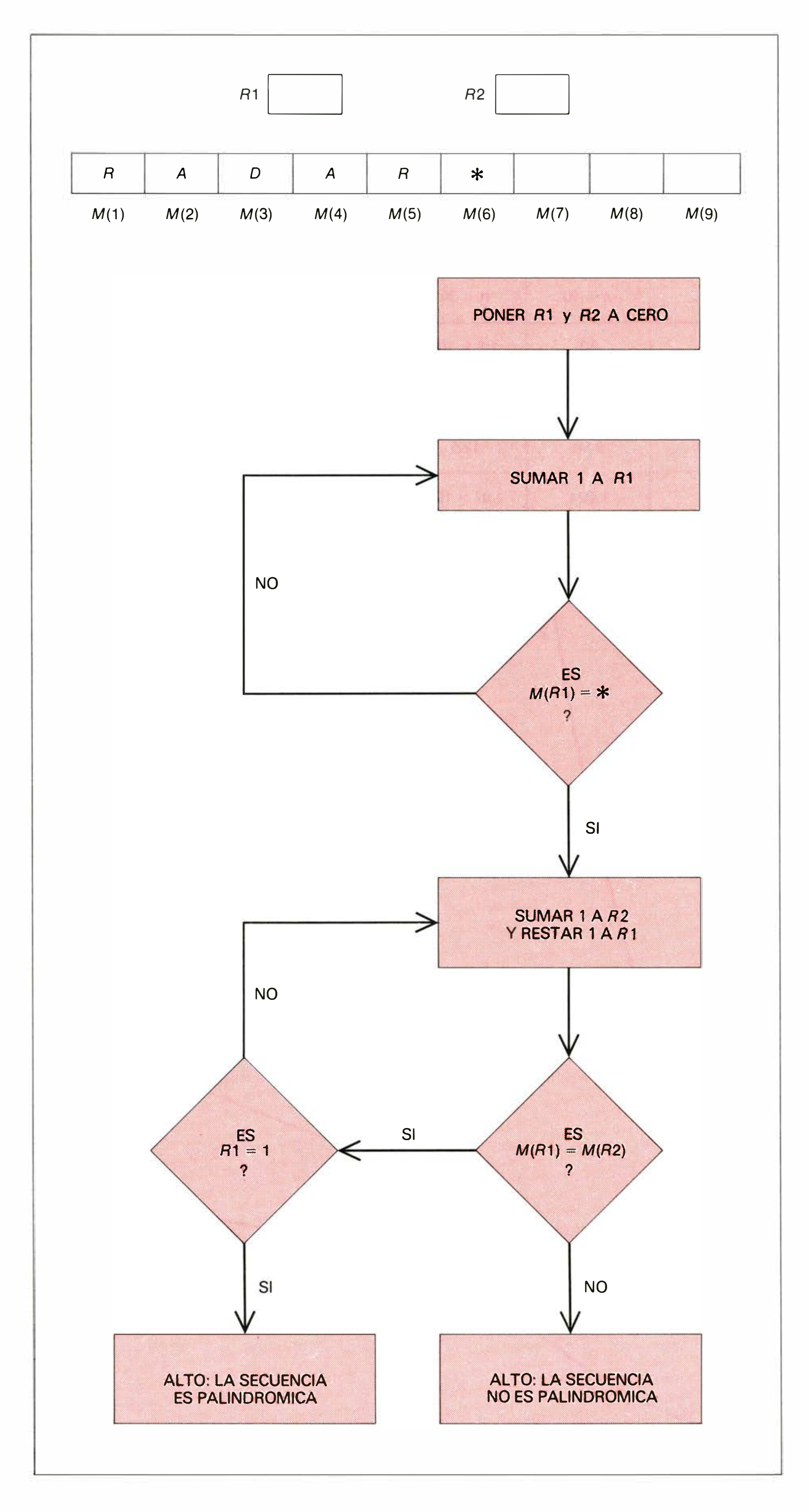
Lo mismo que la máquina de Turing, los modernos ordenadores de gran velocidad tienen una memoria y un sistema de instrucciones para recuperar informaciones archivadas en la memoria, procesarlas y dar entrada en la memoria a nuevas informaciones. Por lo general, la memoria está dividida en una serie de células elementales, a las que se puede acceder por separado. Debido a que estos ordenadores tienen capacidad para direccionar ("address") cualquier célula de memoria, reciben la calificación de máquinas de acceso aleatorio. Está demostrado que una máquina de acceso aleatorio solamente puede resolver aquellos problemas que pueda resolver una máquina de Turing. A pesar de que la máquina de Turing pueda necesitar muchos pasos para simular una sola acción de la máquina de acceso aleatorio, la diferencia en potencia de cómputo se va desvaneciendo al ir adentrándose en tiempos de ejecución exponencialmente crecientes. Si una máquina de acceso aleatorio puede resolver un problema en tiempos con crecimiento exponencial, lo mismo es cierto para una máquina de Turing, y recíprocamente. Cuando al describir un problema se afirma que exige tiempos de resolución c^n independientemente del algoritmo, tal descripción es válida para algoritmos de máquinas de Turing, máquinas de acceso aleatorio y cualesquiera otros computadores que hoy existan o que puedan llegar a construirse en el futuro. No obstante, en general el valor concreto de c sí depende del dispositivo particular de cálculo empleado.

¿De qué forma sería posible demostrar que un problema requiere tiempos de solución con crecimiento exponencial, independientemente del algoritmo empleado para resolverlo? El método más directo sería imaginar todos los posibles algoritmos correspondientes al problema y demostrar después que todos ellos exigen intervalos de tiempo exponenciales. Para la mayoría de los problemas, este método directo no ha sido fructífero. La dificultad está en que es preciso examinar no sólo los métodos evidentes de resolución del problema sino también los más sutiles y profundos. Los principios que inspiran los posibles algoritmos pueden ser sumamente insospechados, y es difícil saber si todos los posibles caminos de resolución del problema han sido analizados y descubiertos. Por ejemplo, las personas no familiarizadas con la

FUNCION	TIPO	$n = 1$	$n = 3$	$n = 10$	$n = 30$	$n = 100$	$n = 300$	$n = 1,000$
$6n$	LINEAL	6	18	60	1.8×10^2	6×10^2	1.8×10^3	6×10^3
n^2	CUADRATICA	1	9	10^2	9×10^2	10^4	9×10^4	10^6
1.1^n	EXPONENCIAL	1.1	1.33	2.59	17.5	13,781	2.62×10^{12}	2.47×10^{41}
2^n	EXPONENCIAL	2	8	1,024	1.07×10^9	1.27×10^{30}	2.04×10^{90}	1.07×10^{301}
10^n	EXPONENCIAL	10	10^3	10^{10}	10^{30}	10^{100}	10^{300}	$10^{1,000}$



EFICIENCIA DE UN ALGORITMO: es función de la forma en que va aumentando su tiempo de ejecución conforme la entrada se va haciendo más y más grande. El ritmo de crecimiento se describe mediante una función matemática. Los algoritmos con tiempos de ejecución exponenciales tienen poca utilidad en la programación de ordenadores, porque dichos tiempos de ejecución crecen explosivamente al aumentar el tamaño del problema. El algoritmo que construye el diagrama de juego de una posición de damas es ineficiente, porque su tiempo de ejecución crece exponencialmente, mientras que el algoritmo de suma para el nim es eficiente, porque su tiempo de ejecución sólo crece en función lineal de la entrada. Aunque una función lineal puede superar a una exponencial para pequeños valores de n, a partir de un cierto n (que depende de las funciones) la exponencial es la mayor.



MEMORIA DE ACCESO ALEATORIO de un ordenador. Consta de una sucesión de células M(1), M(2), M(3)... que son directamente direccionables. En este ejemplo, R1 y R2 son dos registros adicionales de almacenamiento. El algoritmo que aquí se muestra determina si una sucesión de caracteres alfabéticos es o no palindrómica, es decir, si daría la misma lectura de izquierda a derecha que de derecha a izquierda, por ejemplo: "radar" (*arriba*). Las sucesiones se introducen en la memoria, seguidas por un símbolo de asterisco (*), que marca el fin de la sucesión. Si al algoritmo se le entrega una sucesión de n caracteres para que la analice, ejecuta la tarea en $6n + 5$ pasos, como máximo.

teoría de juegos posiblemente no se darían cuenta de que el nim admite un algoritmo muy rápido, y supondrían erróneamente que la solución de ese problema requiere tiempos con crecimiento exponencial.

No obstante, este método directo sí ha prestado buenos servicios a la hora de probar que existen efectivamente problemas de complejidad inherentemente exponencial. La existencia de dichos problemas fue establecida a comienzos de la década de los 60 por Michael O. Rabin, de la Universidad Hebrea de Jerusalén, y Juris Hartmanis y Richard E. Stearns, del General Electric Research Laboratory. La construcción de dichos problemas se apoyaba en la técnica de la diagonalización, inventada por Georg Cantor, matemático alemán de finales del siglo pasado, para demostrar que es imposible establecer una correspondencia biunívoca entre el conjunto de los números reales y el de los números enteros positivos [*véase la ilustración superior de la página siguiente*]. Los problemas inherentemente complejos contruidos mediante el procedimiento diagonal solamente son problemas en el sentido abstracto del término, pues no tienen relación con ningún ente tangible. Sin embargo, sí tienen importante papel a la hora de probar que ciertos problemas más a ras del suelo son inherentemente complejos.

En ocasiones es posible relacionar el grado de complejidad de las soluciones de dos problemas en apariencia totalmente inconexos, demostrando que uno de estos problemas es "eficientemente reducible" al otro. Se dice que un problema P es eficientemente reducible al problema Q cuando una solución eficiente de Q proporcionaría una solución eficiente de P. El procedimiento habitual para demostrar que entre P y Q se verifica una relación de este tipo consiste en hallar para P un algoritmo eficiente que en el curso de sus cálculos busca respuesta que debe facilitarle un algoritmo para Q. Es necesario que el algoritmo para Q dé "gratuitamente" sus respuestas; dicho de otra forma, la determinación de respuestas para Q no debe incrementar el tiempo de ejecución del algoritmo correspondiente a P. Por ejemplo, el problema de hallar el cuadrado de un número puede reducirse eficientemente al de multiplicar dos números, porque si se ha de elevar al cuadrado un número dado, el algoritmo de multiplicación lo reduce a multiplicar el número por sí mismo. No es tan evidente que el problema de multiplicación pueda reducirse eficientemente al de elevación al cuadrado, pero la verdad es

	ALGORITMO 1	ALGORITMO 2	ALGORITMO 3	ALGORITMO 4 ...
ENTRADA 1	SI	NO	NO	SI
ENTRADA 2	SI	NO	NO	SI
ENTRADA 3	NO	SI	NO	SI
ENTRADA 4 ⋮	SI ⋮	NO ⋮	SI ⋮	SI ⋱

	PROBLEMA DIAGONAL
ENTRADA 1	NO
ENTRADA 2	SI
ENTRADA 3	SI
ENTRADA 4 ⋮	NO ⋮

DIAGONALIZACION, proceso conceptual para construir problemas intrínsecamente complejos, es decir, problemas que sólo pueden ser resueltos mediante algoritmos cuyos tiempos de ejecución presenten un crecimiento exponencial. Imaginemos una tabla (*izquierda*) cuya primera fila conste de todos los algoritmos que contestan sí o no a los problemas planteados efectuando, a lo sumo, 2^n pasos, donde n representa el tamaño de la entrada. En la primera columna se han colocado todas las posibles entradas para estos algoritmos. Hay una infinidad de filas y de columnas. La tabla nos muestra las respuestas que los distintos algoritmos darían para cada entrada. Tomando las respuestas a lo largo de la diagonal (*color*) y cambiando cada no por sí y cada sí por no, podemos formar un "problema diagonal" intrínsecamente complejo (*derecha*). Por ejemplo, la respuesta al problema diagonal para la entrada 3 es sí porque la tercera fila tiene un no en la diagonal. Esto asegura que las respuestas al problema diagonal no van a coincidir con las respuestas de ningún problema cuyo tiempo de ejecución sea de 2^n a lo sumo. El problema diagonal sólo puede ser resuelto mediante un algoritmo que consuma tiempos con crecimiento exponencial. Decir que un problema es insoluble significa que ningún algoritmo podrá resolverlo.

que el producto de dos números formados por n cifras puede calcularse en tiempos proporcionales a n aplicando dos veces el algoritmo de elevación al cuadrado, porque $xy = (x + y)^2/4 - (x-y)^2/4$. Aparte de la aplicación por dos veces del algoritmo de elevación, los cálculos de la fórmula ocuparían un tiempo proporcional a n, lo cual se considera eficiente en este contexto, porque no se conoce ningún procedimiento para multiplicar números o elevarlos al cuadrado en tiempos tan pequeños.

Si se sabe que un problema P es eficientemente reducible al problema Q, la complejidad de P no podrá ser mucho mayor que la de Q. Fue este hecho el que nos permitió probar que el problema de análisis correspondiente al "roadblock" es tan complejo que es inherentemente exponencial. Denotemos EXP a la clase de todos los problemas que puedan resolverse en tiempos de crecimiento exponencial. Nosotros hemos conseguido demostrar que todo problema perteneciente a EXP es eficientemente reducible al problema de análisis del "roadblock". En nuestro caso, el método de reducción es algo más complicado que en los problemas de multiplicación y elevación al cuadrado, pero de todas formas, el tiempo requerido para efectuar la reducción es despreciable frente a tiempos con crecimiento exponencial; resultado que muestra que la complejidad de un problema perteneciente a EXP no es mucho mayor que la complejidad del problema de análisis para el "roadblock". Como EXP está formado por todos los problemas que pueden resolverse en tiempos con crecimiento exponencial, contiene en particular los problemas de complejidad intrínsecamente exponencial construidos mediante la técnica de diagonalización. Tales problemas pueden reducirse eficientemente al "roadblock", y en consecuencia la complejidad del problema de análisis del "roadblock" también ha de ser intrínsecamente exponencial.

La clase EXP contiene también el problema de reconocer posiciones ganadoras para el primer jugador en el juego de las damas, y por tanto, este problema es reducible al correspondiente para "roadblock". Cada posición de las damas da lugar a una posición de "roadblock", y ello de tal forma que la posición de damas es ganadora para el primer jugador siempre que la posición en "roadblock" lo sea. Esta correspondencia no sólo puede establecerse con las damas, vale también para el ajedrez, go, hex, nim y otros muchos juegos no goberna-

VARIABLES		CONECTIVAS	CUANTIFICADORES	PERTENENCIA
$a,b,c,\ldots,z$	representan números	$\wedge$ = y	$\forall$ = para todo	$x \in S$ = x pertenece al conjunto S
$A,B,C,\ldots,Z$	representan conjuntos de números	$\vee$ = o	$\exists$ = existe	$x \notin S$ = x no pertenece al conjunto S
		$\rightarrow$ = implica		
		$\leftrightarrow$ = lógicamente equivalente		

ENUNCIADO 1 = $(\forall B)[(\exists x)(x \in B) \rightarrow (\exists s)(s \in B \wedge (\forall y)(y \in B \rightarrow s \leq y))]$

ENUNCIADO 2 = $(\exists D)(\exists E)[0 \notin D \wedge 1 \notin E \wedge (\forall b)(b \notin D \rightarrow b+2 \notin D) \wedge (\forall b)(b \notin E \rightarrow b+2 \notin E) \wedge (\exists t)(t \in D \wedge t \in E)]$

LENGUAJE DE SEGUNDO ORDEN de un sucesor (*S1S*). Se trata de un lenguaje matemático que sirve para expresar con precisión enunciados relativos a números naturales y a conjuntos de números naturales. Damos aquí la lista de símbolos para denotar variables, conectivas, cuantificadores y pertenencia a conjuntos, usada en el lenguaje *S1S*. El enunciado 1 es un aserto verdadero, "todo conjunto no vacío *B* contiene un elemento mínimo"; el enunciado 2 es la falsa aseveración de que "existen conjuntos *D* y *E* y un número *t* tales que *D* está formado solamente por números impares, *E* está formado solamente por números pares, y *t* pertenece a la vez a *D* y a *E*." El problema de decisión para *S1S* consiste en determinar si un enunciado dado es verdadero o falso. Se ha demostrado que el problema de decisión podría ser resuelto mediante algoritmos, aunque cualquiera de ellos tendría que consumir un tiempo de crecimiento superexponencial para decidir la veracidad o falsedad de ciertos enunciados. J. R. Büchi demostró esa posibilidad.

dos por el azar. Por ello el "roadblock" tiene propiedades de juego universal: si pudiéramos resolverlo dispondríamos de soluciones para otros muchos juegos.

La técnica de reducción eficiente ha sido explotada para demostrar teoremas referentes a la complejidad intrínseca de reconocer posiciones ganadoras en el go y en las damas, cuando tales juegos se generalizan a tableros arbitrariamente grandes. El go ha sido estudiado por David Lichtenstein y Michael Sipser, de la Universidad de California en Berkeley; y las damas, por A. S. Fraenkel e Y. Yesha, del Weizmann Institute of Science, en Israel, M. R. Garey y D. S. Johnson, de los Bell Laboratories, y T. Schaefer, de la CALMA Company. Estos autores han demostrado que una solución eficiente del go o de las damas permitiría disponer de soluciones eficientes para otros muchos problemas para los que se han buscado infructuosamente algoritmos rápidos. La posibilidad de que dichos problemas admitan algoritmos rápidos es muy remota, pero no puede todavía descartarse completamente, y por ello la demostración no prueba de forma definitiva que las damas y el go sean intrínsecamente exponenciales, aunque sí da un paso importante en esta dirección. Es posible que el método inventado por los autores para probar la complejidad intrínseca del "roadblock" no funcione en el caso de las damas o del go, y que sea preciso inventar nuevos y más ingeniosos métodos. No obstante, el método de establecer la dificultad intrínseca de un problema por diagonalización y posterior reducción eficiente funciona no sólo para el "roadblock", sino también en ciertos problemas de lógica y de teoría de números.

Hay gran número de conjeturas matemáticas que pueden enunciarse de forma sencilla y perfectamente determinada, y a pesar de ello, su veracidad o falsedad ser un misterio. Quizá la más famosa de estas conjeturas sea el llamado "último teorema de Fermat". El prodigioso matemático francés Pierre de Fermat, considerado padre de la moderna teoría de números, vivió en el siglo XVII. Los márgenes de su ejemplar de un famoso tratado de aritmética griega contienen numerosas notas escritas por él; una de ellas contiene la proposición que ha llegado hasta nuestros días con el nombre de último teorema de Fermat (teorema "magno"): "Es imposible que un cubo sea suma de dos cubos, una cuarta potencia, suma de dos cuartas potencias, o en general, que ningún número que sea potencia mayor que la segunda pueda ser suma de dos potencias semejantes. He descubierto una demostración verdaderamente maravillosa de esta proposición que este margen es demasiado estrecho para contener." Dicho con otras palabras, no existen enteros positivos x, y, z y n, siendo n mayor que 2, tales que $x^n + y^n = z^n$. A pesar de que los más capaces matemáticos de los tres últimos siglos han intentado demostrar (o infirmar) esta proposición, ninguno ha tenido éxito [véase "El último teorema de Fermat", por Harold M. Edwards, en INVESTIGACION Y CIENCIA, diciembre de 1978].

CLASE DE ENUNCIADOS		TIEMPO DE EJECUCION	
ENUNCIADOS QUE CONTIENEN VARIABLES NUMERICAS	¿PUEDE HABER CONJUNTOS?	AL MENOS	A LO SUMO
VARIABLE + NUMERAL	SI	$\left.2^{2^{\cdot^{\cdot^{\cdot^{2^{n}}}}}}\right\} cn$	$\left.2^{2^{\cdot^{\cdot^{\cdot^{2^{n}}}}}}\right\} dn$
VARIABLE + VARIABLE		$2^{2^{cn}}$	$2^{2^{2^{dn}}}$
VARIABLE x VARIABLE		$2^{2^{2^{cn}}}$	$2^{2^{2^{2^{dn}}}}$

COMPLEJIDAD INTRINSECA de los problemas de decisión que se presentan en lógica. Ha sido clasificada para muchos casos. Las dos primeras columnas de esta tabla describen clases de enunciados, y las dos últimas dan las cotas superior e inferior hoy conocidas para los tiempos de ejecución de los algoritmos que resuelven el problema de decisión correspondiente a cada clase. Los tiempos se expresan en función de la longitud del enunciado, *n*; *cF3 y d* son constantes mayores que cero. Los enunciados de la primera fila son del lenguaje *S1S*. Los de la segunda, de la aritmética de Presburger.

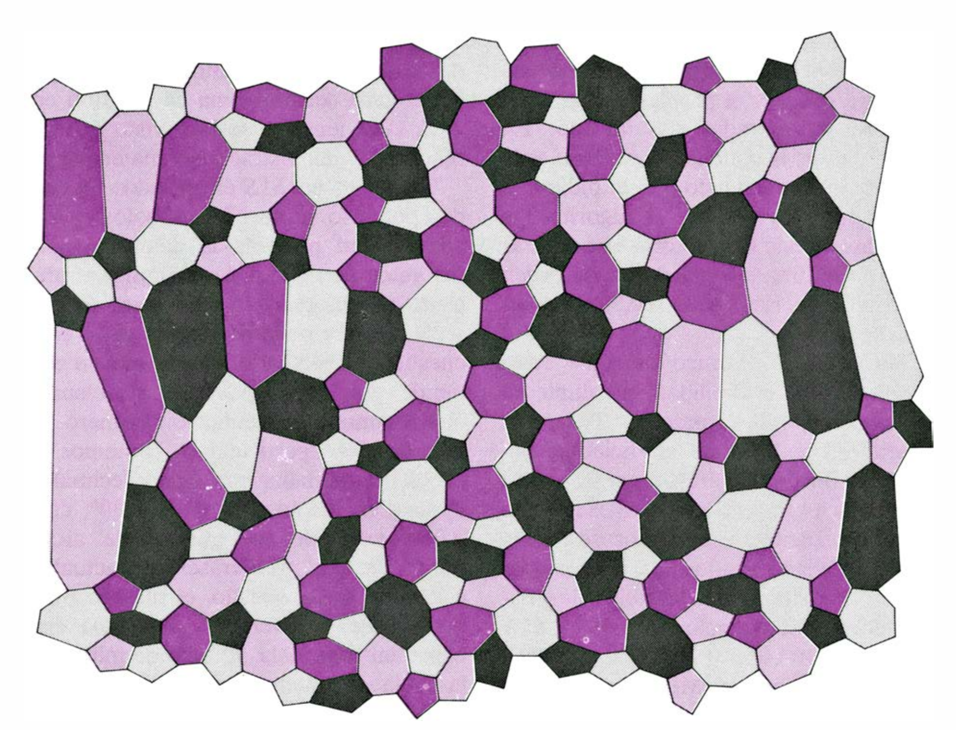
En otros casos, pudo demostrarse la veracidad de la conjetura tras muchos años de eludir los esfuerzos de los mejores matemáticos. Una de ellas es el llamado teorema del mapa de cuatro colores, propuesta en 1852, pero no demostrada hasta 1976, por Kenneth Appel y Wolfgang Haken, ambos de la Universidad de Illinois. El teorema enuncia que es posible pintar todos los países de un mapa plano utilizando solamente cuatro colores, de forma que nunca dos países fronterizos estén pintados del mismo color. Appel y Haken consiguieron su demostración con la colosal ayuda de computadoras de gran velocidad, que emplearon 1200 horas de funcionamiento coloreando unos 2000 mapas de hasta 200.000 formas [véase "La solución del problema del mapa de cuatro colores", por Kenneth Appel y Wolfgang Haken, en INVESTIGACION Y CIENCIA, diciembre de 1977].

En vista del papel fundamental que tuvieron los ordenadores en la demostración del teorema del mapa de cuatro colores, vale la pena examinar hasta qué punto podrían los ordenadores asumir la tarea de dictaminar la veracidad de proposiciones matemáticas. ¿Podría existir un algoritmo tal que al suministrarle enunciados matemáticos redactados en lenguaje suficientemente preciso el algoritmo llegase finalmente a determinar si el enunciado es verdadero o falso? La respuesta es negativa. Se ha demostrado que un tal algoritmo es imposible en el caso de un potente lenguaje formalizado llamado cálculo de predicados. Los problemas no reductibles a solución algorítmica se denominan "insolubles".

A pesar de ello, tan atractiva es la idea de que un ordenador pueda demostrar teoremas de forma automática que su imposibilidad general no ha conseguido disuadir a los matemáticos de intentar materializarla, al menos en parte. Un procedimiento para evitar el escollo de la insolubilidad consiste en restringir la categoría de enunciados propuestos al ordenador. Restricciones que no deben ser tan rígidas que únicamente sobrevivan a ellas enunciados triviales, pero sí

lo suficientemente enérgicas como para erradicar la plaga de insolubilidad. Diversas restricciones de este tipo han sido analizadas por los lógicos. Si el ordenador es limitado a considerar únicamente un cierto tipo de enunciados, entonces, con ayuda de un algoritmo puede decidir correctamente el valor de verdad de cualquiera de esos enunciados. Dicho de otra forma, el problema de decidibilidad es resoluble para este tipo restringido de enunciados.

Ya es hora de examinar una clase concreta de enunciados, correspondiente al lenguaje de segundo orden de un sucesor (*S*1*S*) que es accesible a solución algorítmica. Este lenguaje se refiere a los números naturales (0, 1, 2, 3,...). Los enunciados de máxima sencillez de *S*1*S* son asertos sobre el tamaño relativo de dos números, o sobre la suma de números. Un ejemplo de enunciado verdadero es "3 es menor que 4", y un ejemplo de enunciado falso, "2 + 2 = 5". Las sentencias de este tipo pueden combinarse mediante las conectivas lógicas "y", "o" e "implica". Así, los ejemplos anteriores pueden conectarse mediante "o", dando la proposición " '3 es menor que 4' o '2 + 2 = 5' ". También son lícitos enunciados que mencionan conjuntos de números, como, por ejemplo, "Los números del conjunto (0, 5, 7) son menores que los números del conjunto (9, 12, 15, 19, 24, 100)".



MAPA DE 120 PAISES que pone de manifiesto el significado del teorema del mapa de cuatro colores, el cual afirma que todo mapa plano puede pintarse de forma que nunca dos países fronterizos tengan el mismo color, usando solamente cuatro tintas. El teorema fue conjeturado ya en 1852, pero no pudo ser demostrado hasta 1976 (por Kenneth Appel y Wolfgang Haken, de la Universidad de Illinois). La demostración fue insólitamente larga, y dependió crucialmente de la ayuda de ordenadores de alta velocidad, que invirtieron 1200 horas de funcionamiento coloreando de hasta 200.000 maneras cada mapa de un conjunto de casi 2000. (En *Investigación y Ciencia*, diciembre de 1977, los autores publicaron un extenso artículo sobre el proceso que les condujo a tan feliz descubrimiento topológico.)

La verdadera potencia de *S*1*S* reside en que consiente expresar generalidades matemáticas a través de variables alfabéticas como x, y, z, que representan números, y de cuantificadores, tales como "para todo", "cada" y "existe". Por ejemplo, las sentencias "Cada conjunto no vacío de números tiene un número mínimo" y "Para cada número x existe un número y tal que $y = x + 1$". Para que el valor de verdad de tales enunciados pueda ser determinado mediante un algoritmo, es necesario limitar *S*1*S* a enunciados que no contengan la suma de dos variables.

En 1960, J. R. Büchi, de la Universidad de Michigan, demostró que existe un algoritmo capaz de determinar correctamente el valor de verdad de cualquier enunciado dado perteneciente a *S*1*S*. Resulta algo sorprendente que *S*1*S* sea resoluble, en vista del carácter "abierto" de los cuantificadores. Por ejemplo, para determinar el valor de verdad del enunciado "para todo x, 2 es menor que $x + 3$", el algoritmo no podría intentar la sustitución de x por cada uno de los números naturales, porque éstos forman un conjunto infinito. En vez de ello, el algoritmo tendrá que valerse de algún otro tipo de razonamientos; tal vez, efectuar ciertos cálculos algebraicos y aplicar el principio de inducción completa. En efecto, Büchi ha demostrado que existen algoritmos capaces de llevar a término los razonamientos necesarios para establecer el valor de verdad de cualquier enunciado de *S*1*S*, independientemente de la longitud del enunciado.

Los autores pudieron demostrar que el problema de decisión para el "roadblock" admite solución algorítmica, pero tal demostración nada dice sobre la eficiencia del algoritmo, que ha resultado ser intrínsecamente exponencial. De la misma forma, la demostración de existencia de solución algorítmica para el problema de decisión en *S*1*S*, dada por Büchi, nada promete sobre la eficiencia del algoritmo. De hecho, Meyer demostró en 1972 que la complejidad intrínseca del problema de decisión para *S*1*S* crece muchísimo más rápidamente que las funciones exponenciales. Por ejemplo, las funciones doblemente exponenciales, como 2^{2^n} manifiestan crecimientos super-exponenciales. Para calcular el de una función doblemente exponencial para un cierto número n se empieza por exponenciar este número, dando así un exponente intermedio $m = 2^n$, resultado que a su vez vuelve a utilizarse como exponente, con lo que se obtiene el resultado final 2^m. Análogamente, en una función triplemente exponencial se efectúan tres exponenciaciones sucesivas; en una exponencial cuádruple, cuatro exponenciaciones consecutivas, y en general, dado un número k natural, iterando k veces el procedimiento de exponenciación se obtiene una función k-múltiplemente exponencial. Es difícil imaginar la velocísima tasa de crecimiento de las funciones múltiplemente exponenciales, pues ningún proceso físico conocido exhibe un comportamiento tan explosivo.

El notable teorema de Meyer afirma que, por grande que sea el número k de sucesivas exponenciaciones que se tome, la complejidad intrínseca del problema de establecer el valor de verdad de los enunciados de *S*1*S* aumenta más rápidamente que las funciones k-múltiplemente exponenciales de la longitud n del enunciado. Si bien el teorema de Büchi afirma que el problema de decidibilidad para *S*1*S* es en principio resoluble, el teorema de Meyer viene a tener sensiblemente las mismas consecuencias que una demostración de insolubilidad. Decir que un problema es insoluble significa que ningún algoritmo podrá resolver el problema para un número infinito de datos de entrada. Gran complejidad

intrínseca significa que todo algoritmo requerirá tiempos intolerablemente largos para resolver el problema cuando el número de entradas sea infinito. En ambos casos, la conclusión práctica es la misma: un matemático que le pidiese al ordenador que aplique el algoritmo a entradas de este tipo tendría que esperar demasiado (una eternidad, sea de orden práctico o teórico) para recibir la solución.

No es *S*1*S* el único lenguaje cuyo problema de decidibilidad es soluble en teoría pero no en la práctica. Tomemos ahora el conjunto de enunciados, llamado aritmética de Presburger, formado por todos los enunciados de *S*1*S* excepto los que mencionan conjuntos de números, y ampliado con los enunciados relativos a las sumas de variables numéricas, como por ejemplo "$x + y = z$" y "$x + z = 30$". En 1973, Rabin y Michael J. Fischer, del MIT, demostraron que el problema de decisión para la aritmética de Presburger tiene complejidad intrínseca de crecimiento al menos tan rápido como una función doblemente exponencial de la longitud de enunciado. En años más recientes se ha descubierto que los problemas de decisión de otros muchos lenguajes manifiestan también complejidad super-exponencial. Tales descubrimientos revelan que la pretensión de evitar que el problema de decidibilidad de enunciados lógicos resulte insoluble, restringiendo a tal fin las categorías de enunciados admisibles, tropieza con muy serios obstáculos.

Otra posible vía de eludir la plaga de insolubilidad se apoya en la suposición de que en las aplicaciones reales, nadie está interesado en resolver todos los casos particulares que pueda ofrecer un problema, sino que tan sólo se desea conocer la solución de un número finito de ellos. Por ejemplo, es improbable que nadie desee conocer la veracidad o falsedad de enunciados compuestos por un millón de símbolos. Los resultados sobre complejidad intrínseca que hasta ahora hemos descrito han venido refiriéndose a los desmesurados crecimientos de los tiempos de ejecución, en función del tamaño de las entradas. Hemos precisado bastante al hablar de tales crecimientos (exponencial, doblemente exponencial, triplemente exponencial, etc.), pero en cambio hemos detallado muy poco el momento en que el tamaño de las entradas hace que el crecimiento explosivo del tiempo de ejecución resulte inadmisible. Podría sospecharse que si la complejidad intrínseca de un problema creciera exponencialmente cuando menos, el problema continuaría siendo demasiado complejo, incluso reduciendo las entradas a unos pocos cientos de símbolos. En el caso del problema de decisión en *S*1*S*, tales temores se han visto confirmados por demostraciones matemáticas

Tomemos en *S*1*S* enunciados formados por a lo sumo 675 símbolos. ¿Qué complejidad presenta la determinación del valor de verdad de cualquier enunciado perteneciente a esta clase finita? Antes de que podamos responder a esta cuestión es necesario aclarar qué se entiende por complejidad de problemas que solamente presentan un número finito de casos particulares. Tomemos el problema de hallar con cinco decimales el seno de los ángulos de 0 a 90°, con incrementos de un minuto de arco. Antes de que proliferasen las actuales calculadoras de bolsillo, el método más normal de resolver este problema era consultar una tabla. El acto de mirar en la tabla es mucho más rápido que el cálculo del seno utilizando una fórmula y haciendo operaciones con lápiz y papel. El aumento de velocidad se paga con el aumento de tamaño del algoritmo: la tabla trigonométrica ocupa mucho más espacio que la fórmula.

Una vez construida una tabla de soluciones de un problema finito, a base de resolver y registrar la respuesta de cada una de las posibles entradas, el método de resolver el problema mirando la tabla tiene todas las características de los algoritmos, pudiendo aplicarse de forma automática a cada caso concreto del problema. La pega está en que el número de casos de un problema finito puede ser tan grande que resulte imposible de manejar. En *S*1*S* hay considerablemente más de 10^{100} enunciados de longitud 675. Es evidente que la construcción de una tabla que conste de una entrada para cada enunciado es impensable. La idea que hemos querido dejar clara es que al hablar de la complejidad de problemas finitos es necesario tener en cuenta no sólo el tiempo de ejecución del algoritmo, sino además, su propio tamaño.

Un método conceptualmente sencillo de medir el tamaño de un algoritmo consiste en contar el número de componentes que habría de tener un computador capaz de ejecutarlo. Se supone que el ordenador está diseñado para resolver exclusivamente este problema, y que no hay, por tanto, circuitos redundantes. En los ordenadores reales, las unidades básicas de lógica electrónica son las "puertas" (gates), o conmutadores elementales. Cada puerta toma dos señales eléctricas, cada una de las cuales puede ser de dos tipos, que llamaremos "estado alto" (conmutado, on) y "estado bajo" (desconectado, off), y entrega en su salida una tercera señal, de tipo igual que las de entrada. Estas puertas son los componentes electrónicos más elementales, pero interconectando miles de ellas llegan a construirse sistemas muy complejos. Cuando una de las señales que llegan a la puerta cambia de estado alto a estado bajo, o viceversa, se produce una breve demora hasta que la salida alcanza un valor estable. El tiempo que el ordenador necesita para ejecutar el algoritmo es, como mínimo, el producto del retardo de conmutación multiplicado por el número de veces que las puertas deben cambiar de estado.

Los ordenadores rápidos de mayor tamaño tienen alrededor de medio millón de puertas, con retardos de aproximadamente 10^{-9} segundos en cada conmutación. Evidentemente, se irán construyendo computadores cada vez más grandes y más rápidos, pero ninguno podrá jamás superar la capacidad del máximo ordenador imaginable, mencionado al principio del artículo. Este ordenador ideal estaría formado por unas 10^{126} puertas no menores que un protón. Es razonable suponer que el retardo de conmutación no podrá ser inferior a los 10^{-24} segundos que la luz necesitaría para atravesar el diámetro del protón. Meyer y uno de nosotros (Stockmeyer) demostraron que este imaginario ordenador precisaría al menos 20.000 millones de años para determinar el valor de verdad de ciertos enunciados de *S*1*S* formados por 675 símbolos, lo que significa que hay enunciados de longitud 675 cuyo valor de verdad nunca podrá ser determinado mediante algoritmos.

Es posible que estos inescrutables enunciados sean muy raros y escasos. Además, puede que para efectos prácticos no sea necesario esperar de un algoritmo que pueda manejar todos los enunciados de hasta una cierta longitud. En el caso del *S*1*S* podríamos imaginar un algoritmo que diese una de estas tres respuestas: "verdadero", "falso" y "no lo sé". La posibilidad de que se presentasen respuestas "no lo sé" no menoscabaría demasiado la utilidad del algoritmo si solamente ocurriesen en raras ocasiones. Aunque puede parecer que este algoritmo "trivalente" torea limpiamente la no computabilidad del *S*1*S*, no hay ninguna garantía de que sea algoritmo eficiente. La cuestión de si la posibilidad de respuestas ocasionales indeterminadas hará más sencillo el problema de decisión en *S*1*S* es cuestión que todavía permanece abierta.

El cometa de Halley pintado por Giotto

Un cometa resplandeciente representa la estrella de Belén en un famoso fresco de Giotto. Es un retrato naturalista del cometa de Halley tal como pudo observarse durante su aparición espectacular de 1301

Roberta J. M. Olson

A finales del verano de 1301 un cometa espectacular cruzó la cúpula del firmamento nocturno. Su luminosa cabellera esférica rodeaba una incandescencia central, similar a una estrella, que parecía arrojar un fulgor difuso de partículas radiantes. Sabemos ahora que se trataba del cometa de Halley en una de sus visitas periódicas al centro del sistema solar. Sabemos también cómo apareció ante los ojos asombrados de los observadores italianos porque su "retrato" fue realizado con cuidadoso detalle por Giotto di Bondone (1267-1337), el pionero florentino de la pintura naturalista.

Giotto estaba en Italia en 1301 (aunque no sabemos con seguridad en qué lugar preciso se encontraba) y muy probablemente debió ver el cometa. En un plazo comprendido entre uno y cuatro años realizó la extraordinaria serie de frescos de los muros de la Capilla Scrovegni (Arena) en Padua, en el norte de Italia. En una escena que representa la Adoración de los Magos representó la estrella de Belén: en ella abandonó valientemente la tradición iconográfica al no dibujarla como una pequeña estrella estilizada con muchas puntas, sino como un cometa resplandeciente. La coincidencia de las fechas, la representación naturalista del cometa y su semejanza con las fotografías del mismo realizadas a raíz de su aparición más reciente, en 1910, muestran claramente que el cometa de Giotto es, en realidad, el de Halley.

Existen miles de millones de cometas (la palabra deriva del griego *kometes,* "de larga cabellera") que describen su órbita en torno al sol en una nube tenuemente distribuida que llega millares de veces más lejos que los planetas más alejados. El núcleo de cada cometa es un pequeño conglomerado gélido de gases congelados y polvo interestelar. A veces una de estas "sucias bolas de nieve" cuya órbita ha sido, tal vez, perturbada por una estrella, penetra en las regiones interiores del sistema solar donde sus gases y polvo, activados por la radiación y el viento solares, forman una cabellera visible o conjunto formado por cabeza y cola. Algunos de estos cometas tienen una órbita elíptica que, de manera periódica, les hace visibles desde la tierra.

El cometa de Halley fue el primero de estos cuerpos celestes cuya periodicidad fue reconocida. En 1705 el astrónomo inglés Edmund Halley señaló la existencia de notables semejanzas en las órbitas que había calculado, basadas en observaciones anteriores, para las apariciones de grandes cometas en 1682, 1607 y 1531. Postuló que las tres apariciones representaban visitas repetidas, separadas por un intervalo aproximado de 76 años, de un único cometa que describía una órbita elíptica alargada en torno al sol. Predijo que el cometa regresaría en 1758. Los astrónomos posteriores, que tuvieron en cuenta las perturbaciones introducidas en la órbita por Júpiter y otros planetas, introdujeron mejoras en los cálculos de Halley y predijeron que el próximo paso por el perihelio (posición más próxima al sol) del cometa de Halley tendría lugar en abril de 1759. El cometa fue observado por vez primera por un campesino el día de Navidad de 1758 y alcanzó su perihelio en marzo de 1759.

Otros cálculos posteriores han establecido la fecha de cada paso por el perihelio hasta el 239 a. de C.; el estudio de los testimonios antiguos (sobre todo los de los observadores chinos para pasos anteriores al siglo XV) revela la existencia de observaciones del cometa en el momento adecuado y en el lugar apropiado del cielo que coinciden con la mayoría de los retornos al perihelio calculados. El período medio de revolución del cometa es algo inferior a 77 años, pero el período real llega a variar hasta dos años y medio por encima y por debajo de la media debido a perturbaciones introducidas por los planetas. El cometa de Halley volverá a aparecer en 1985 y se ha calculado que el paso por el perihelio tendrá lugar el 9 de febrero de 1986.

Se estima que, en cada siglo, un número comprendido, quizás, entre 20 y 30 cometas resultan visibles desde la tierra. En el siglo XIII, dos siglos antes de la invención del telescopio, se registraron más de 20 observaciones de "cometas" realizadas a simple vista. La mayor parte de estos "cometas" a los que se refieren los cronistas occidentales son objeto de una única referencia que carece de precisión: algunos de ellos fueron, sin duda, meteoros y otros fenómenos celestes. Las únicas apariciones de cometas que resultan dignas de confianza son las que han sido verificadas por una correlación cuidadosa de varias observaciones que incluyen las que aparecen en los anales chinos, introducidos en Occidente por los misioneros jesuitas que regresaron en el siglo XVIII, y traducidos por vez primera en 1846. Estas debieron ser, probablemente, las apariciones más espectaculares ya que atrajeron suficiente atención como para hacer surgir referencias múltiples en Europa y en Asia. Un meteoro podía ser registrado como un cometa por una sola persona pero un cometa auténtico no sería confundido con facilidad por varios testigos. Una aparicion espectacular asombraría a todos los observadores.

De entre los cometas cuyas órbitas calculadas coinciden con los registros que aparecen en las crónicas del siglo XIII, sólo dos pueden considerarse apariciones espectaculares. El primero

fue observado en China y Europa en 1264. El segundo fue, sin duda, visto en China en 1299 (y un tratado atribuido a un tal Pedro de Limoges registra una observación de este cometa en Europa). Giotto nació hacia 1267, y, por tanto, no pudo contemplar el cometa de 1264. La aparición del cometa de 1299 no se debió ver, probablemente, en Italia o bien su observación resultó poco conspicua allí. La primera aparición espectacular registrada con seguridad dentro de la vida de Giotto es, por consiguiente, la del cometa de Halley en 1301. La siguiente aparición espectacular de un cometa no se produjo hasta 1337, el año de la muerte de Giotto. Por otra parte la aparición de 1301 debió resultar impresionante y fue objeto de comentarios por parte de muchos historiadores contemporáneos. El eminente cronista florentino del siglo XIV Giovanni Villani escribió en sus *Chroniche Storiche* que un cometa apareció en los cielos en septiembre de este año "dejando tras de sí grandes rastros de humo" y permaneció visible hasta enero de 1302. Existen varias discrepancias en la fecha y algunos autores afirman que la aparición duró sólo seis semanas, pero todos están de acuerdo en que la cola del cometa era impresionantemente larga; se estima que subtendía un arco de hasta 70 grados en el cielo.



LA ADORACION DE LOS MAGOS es una escena de un ciclo de frescos realizado por el maestro florentino Giotto di Bondone. Este ciclo decora el interior de la Capilla de la Arena y le fue encargado por el comerciante de Padua Enrico Scrovegni (tal vez para expiar los pecados de su padre, identificado por Dante en el *Infierno* con el archiusurero). Scrovegni obtuvo el permiso para erigir el edificio en 1302, el lugar fue consagrado en 1303 y los frescos parecen haberse empezado en este mismo año; la *Adoración*, situada en la segunda fila de un ciclo que tenía muchas, debió terminarse en 1304. La escena ejemplifica las principales innovaciones de Giotto, su naturalismo y la humanidad de sus figuras, y es notable por su representación de la estrella de Belén, no como una estrella estilizada sino como un cometa dinámico. El cometa de Halley, que regresa al centro del sistema solar aproximadamente cada 77 años, hizo una aparición en 1301 y sirvió de modelo al cometa de Giotto; volverá a aparecer, de nuevo en 1985.



UN PRIMER PLANO DEL COMETA en la *Adoración* muestra cómo el artista aplicaba el temple y pigmentos dorados a la pared enyesada con pinceladas que construían una textura con la que intentaba aproximarse a la apariencia luminosa de la cabellera y de la cola del cometa real que Giotto debió observar con cuidado unos pocos años antes. Representó el centro intensamente brillante de la cabellera, o cabeza, con lo que parece ser una estrella de ocho puntas, acumulando capas de color sobre la estrella para que la imagen resultara más difusa. Parte de la pigmentación se ha perdido y queda patente el material adhesivo de color rojo con el que se aplicaba al yeso. (Un cometa describe su órbita entorno al sol en una nube.)

Tal aparición coincide muy bien con la representación de Giotto.

A lo largo de la historia la aparición de un cometa ha sido siempre motivo de miedo. Durante siglos se creyó que todos los aspectos de la vida sobre la tierra estaban gobernados por las posiciones de las estrellas y de los planetas. Una vez se hubieron conocido la situación de las estrellas fijas y las posiciones cambiantes de los planetas, la aparición de una "nueva estrella" de cualquier clase parecía violar el orden de los cielos y, por consiguiente, constituir un indicio de catástrofe (una "mala estrella" en sentido literal). Tales apariciones se asociaban con frecuencia con acontecimientos humanos significativos, pero antes del siglo XVI es muy raro que los cometas sean objeto de representación en el arte occidental (un ejemplo es un cometa en forma de espada que aparece suspendido en el firmamento en un camafeo romano). En la Edad Media los cometas surgen a veces en posiciones secundarias en grandes obras astrológicas, tales como el ciclo zodiacal del siglo XII en la fachada occidental de la catedral de Piacenza en la Italia central. Tales representaciones de los cometas suelen, no obstante, ser genéricas; "retratos" de cometas identificables –representaciones explícitas de apariciones históricas– son aún más raros.

Se han identificado tres de estos "retratos" de apariciones del cometa de Halley. La más antigua representa la aparición correspondiente al año 684 de nuestra era, aunque no fue realizada hasta ocho siglos después del suceso, en el *Liber Chronicorum* o *Weltchronik* de Hartmann Schedel, conocido también como las *Crónicas de Nuremberg* porque fue editado en esta ciudad. El libro fue publicado por primera vez en 1493 y fue ilustrado con grabados en madera debidos al artista alemán Michel Wohlgemuth y a su hijastro Wilhelm Pleydenwurff. El impresor fue Anton Koberger, el padrino de Albrecht Dürer (quien, más tarde, representaría la aparición espectacular de un cometa en su grabado *Melencolia I*). Una representación burda del cometa de Halley aparece, acompañada de un texto ilustrativo, en la página relativa al año 684. (Esta aparece repetida al azar a lo largo del libro porque el limitado repertorio de pequeños grabados en madera tenía una doble función: servir de ilustración especifica y constituir una serie de puntos de referencia para el lector que buscaba un pasaje determinado en el gran volumen en el que las páginas carecen de numeración.)

El más antiguo retrato contemporáneo, aunque resulte estilizado, del cometa de Halley, representa su aparición en la primavera de 1066. Se encuentra en una escena del tapiz de Bayeux que fue encargado por la Reina Matilda, la esposa de Guillermo el Conquistador, para ilustrar la victoria de su marido en la batalla de Hastings. El tapiz (de hecho un bordado en estambre sobre ocho piezas alargadas de lino tosco que miden unos 70 metros de largo por medio metro de ancho) fue realizado entre 1073 y 1083 y ahora se encuentra en el Ayuntamiento de Bayeux en Normandía. Numerosos relatos de los cronistas contemporáneos fechan la aparición de 1066 y señalan su naturaleza espectacular; de ello da testimonio el tapiz mismo, si nos fijamos en la cara de asombro de los espectadores y en la leyenda bordada: "Sienten miedo de la estrella". El cometa está representado de acuerdo con el estilo románico predominante en la época, con una estructura geométrica altamente estilizada y una configuración plana y puramente decorativa constituida por líneas y planos.

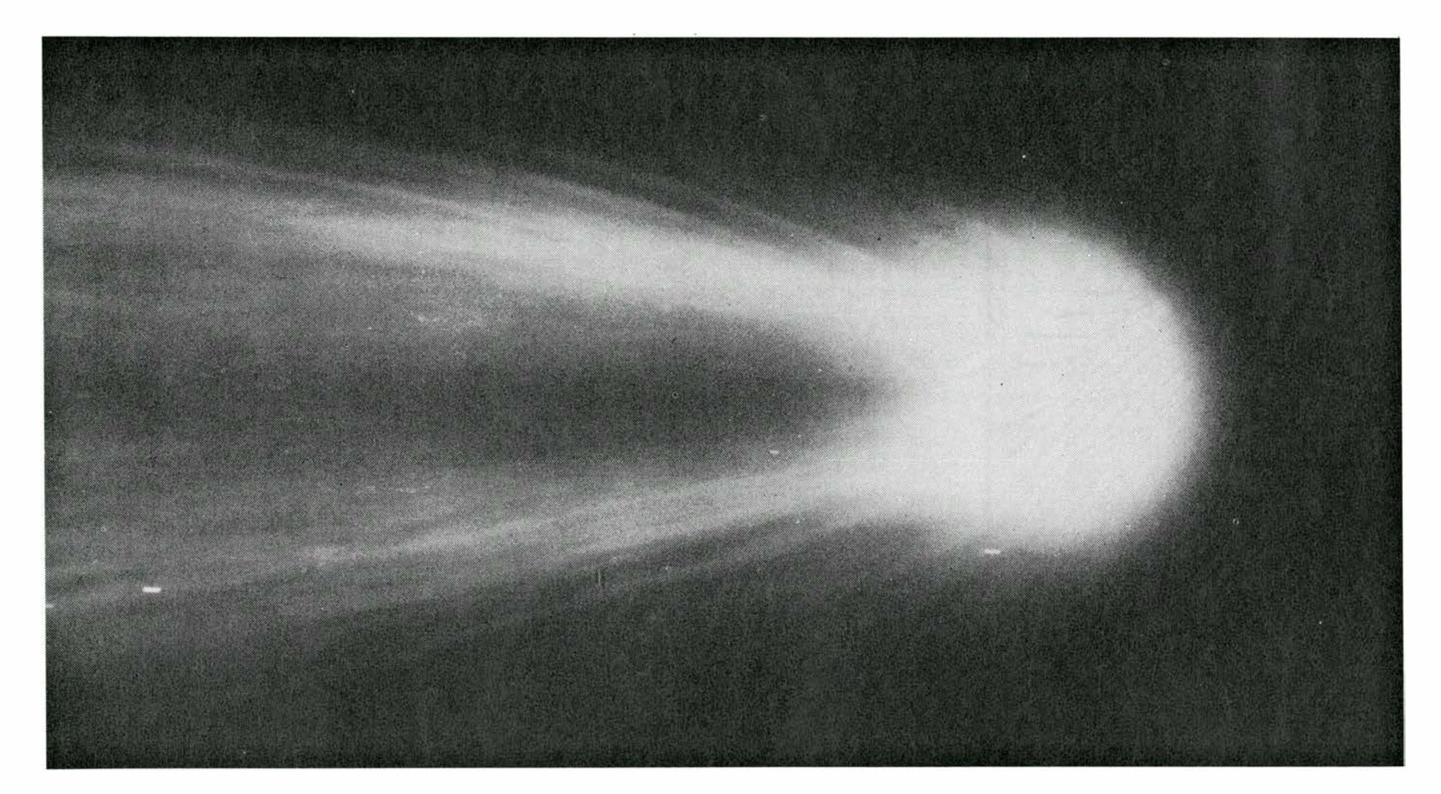
El retrato siguiente del cometa de Halley se encuentra en el *Salterio Eadwine (Canterbury)*, un manuscrito inglés del siglo XII que es una copia del *Salterio de Utrecht*, una colección de los salmos. En la parte inferior de la página en la que se encuentra el texto del salmo 5 aparece una sencilla silueta del cometa: una roseta estilizada inscrita en un círculo, con una cola de cuatro rayos ondulados. La posición del boceto en la parte inferior de la página y su gran escala sugieren que fue añadido, junto con la descripción escrita en anglosajón, como nota incidental; aparentemente carece de relación con las tres versiones latinas del salmo que se encuentran en la página. El estilo del manuscrito y las fechas que corresponden a su escriba, el monje Eadwine, coinciden con la aparición del cometa de Halley en 1145. Pese a tratarse de un simple boceto, este dibujo es revolucionario para el siglo XII

ya que se trata de una representación contemporánea de un fenómeno natural.

El aspecto más notable del retrato que Giotto hizo del cometa de Halley, en violento contraste con las representaciones más antiguas, es su naturalismo. Nada tiene esto de particular ya que la reputación contemporánea de Giotto y su inmensa significación para la historia de la pintura (fue el primer artista que ejerció una influencia casi universal en la pintura occidental) procede de sus sorprendentes innovaciones naturalistas. Nació en el pueblecito de Colle di Vespignano, cerca de Florencia. Se sabe poco acerca de la primera parte de su vida pero se cree que fue discípulo del pintor florentino Cimabue, un maestro progresista que empezó a romper con el estilo italo-bizantino predominante. Giotto se rebeló de manera más dramática en contra del estilo establecido y desarrolló una nueva plástica tridimensional que debió ser considerada más abrumadora en su época que lo que aparece, hoy, a los ojos de los peritos actuales. Sus figuras esculturales y majestuosas, dispuestas con calma de modo que parezcan simples masas de color resplandeciente, comunican una intensidad emocional que carece de precedentes en la pintura europea. Fue precisamente en los frescos de la capilla de Padua donde el estilo revolucionario de Giotto alcanzó no sólo la madurez sino también su manifestación más lúcida.

Giotto había trabajado en Roma hacia finales de siglo y se cree que estuvo en Padua en 1302, pintando un fresco que no se conserva. Un rico mercader de Padua, Enrico Scrovegni, le encargó la decoración del interior de una pequeña capilla familiar que estaba en construcción junto al palacio de la familia y cerca de las ruinas de un anfiteatro romano (de donde deriva la apelación usual de Capilla de la Arena). Se ha sugerido que Giotto fuera el arquitecto de la capilla y, en este caso, la fecha de la concepción de los frescos resultaría incluso más próxima a la aparición del cometa: en cualquier caso los frescos parecen haberse iniciado como máximo en 1303. Giotto cubrió los muros con apretadas hileras de escenas que ilustraban la vida de Cristo, desde los acontecimientos que precedieron a su nacimiento (vidas de S. Joaquín y Sta. Ana, padres de la Virgen María y vida de la Virgen) hasta la infancia de Cristo, su vida adulta, su pasión, crucifixión y resurrección, así como el Juicio Final, todo ello dentro de un amplio programa simbólico. La mayor parte de las 38 escenas narrativas, incluyendo la *Adoración,* miden aproximadamente 200 por 185 centímetros, con figuras de tamaño equivalente a la mitad del natural. Dada su posición en la segunda fila, la *Adoración* puede fecharse en 1303 o 1304.

Cuando Giotto procedió a pintar la estrella de Belén, rechazó los estrechos moldes tanto del simbolismo astrológico como de la convención medieval y reprodujo el cometa tal como lo había visto en realidad pocos años antes, cuando iluminaba de manera teatral el cielo nocturno de Italia. El gran cometa ígneo domina el firmamento de su fresco. La cabellera late con energía: en su centro se encuentra la figura de una estrella de aristas agudas que representa el brillante "centro de condensación" que se ve a menudo en el interior de la cabellera más difusa. (No se trata del núcleo que es demasiado pequeño para ser visible en el interior de la cabellera.) La cola estriada produce una impresión dinámica del arco trazado por el paso del cometa a través del cielo. Los cometas espectaculares tienen este mismo aspecto cuando se observan a simple vista y el cometa de Halley debió aparecer de este modo ante los ojos de Giotto. Puede conjeturarse que tal vez dibujó sus observaciones de 1301 y que utilizó más tarde este esquema como modelo (aunque no conservamos ningún dibujo

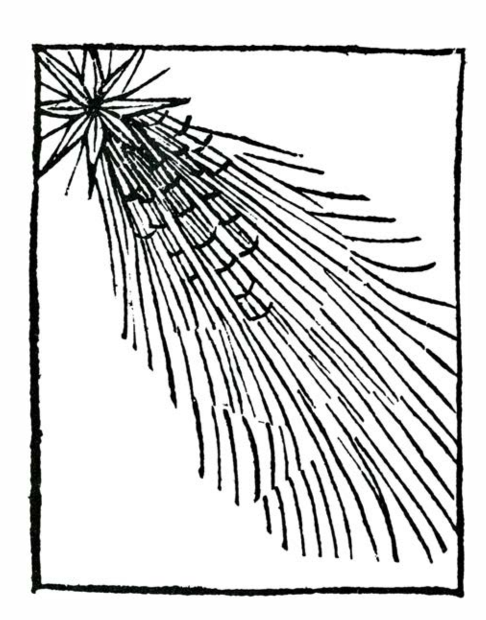


CABEZA DEL COMETA DE HALLEY fotografiada con el telescopio de reflexión de 1524 milímetros de Mount Wilson el 8 de mayo de 1910 durante el retorno más reciente del cometa a las proximidades de la tierra. Los cometas principales tienen características que los distinguen, aunque sean un tanto variables, y esta fotografía resulta reconociblemente similar al retrato de Giotto. El cometa propiamente dicho es un pequeño (y, por tanto, invisible) núcleo de gases congelados y polvo. Cuando el cometa pasa suficientemente cerca del sol, parte del gas se evapora y se lleva consigo partículas de polvo; el brillo de la cabellera del cometa se debe a la luz solar que es dispersada por el polvo o irradiada de nuevo por los gases fluorescentes. La cola, que siempre apunta en la dirección opuesta al sol, es una corriente de polvo impelida por la radiación solar y por electrones que se encuentran en el viento solar e impelida por campos magnéticos situados en el viento solar. El cometa de Halley regresará a su perihelio en 1986.

RETORNOS AL PERIHELIO DEL COMETA DE HALLEY	
239 a.C.	30 MARZO
163	5 OCTUBRE
86	2 AGOSTO
11	5 OCTUBRE
66 d.C.	26 ENERO
141	20 MARZO
218	17 MAYO
295	20 ABRIL
374	16 FEBRERO
451	24 JUNIO
530	25 SEPTIEMBRE
607	13 MARZO
684	28 SEPTIEMBRE
760	22 MAYO
837	27 FEBRERO
912	9 JULIO
989	9 SEPTIEMBRE
1066	23 MARZO
1145	22 ABRIL
1222	1 OCTUBRE
1301	23 OCTUBRE
1378	9 NOVIEMBRE
1456	9 JUNIO
1531	25 AGOSTO
1607	27 OCTUBRE
1682	15 SEPTIEMBRE
1759	13 MARZO
1835	16 NOVIEMBRE
1910	20 ABRIL

REGRESOS AL PERIHELIO, posición más próxima al sol. Tales regresos han sido calculados retrospectivamente hasta el 239 a. de C., y la mayor parte de estas fechas han sido cotejadas con relatos de observaciones. El período medio de la órbita elíptica alargada del cometa es de casi 77 años, pero los períodos sucesivos varían ya que la órbita se halla perturbada por el efecto gravitatorio de los planetas y especialmente de Júpiter.

Cometa



CRONICAS DE NUREMBERG, publicadas en 1493. De ellas hemos sacado esta ilustración, grabada en madera, de un cometa en la página que refiere los acontecimientos del año 684 de nuestra era, un año en el que apareció el cometa de Halley. El texto que la acompaña refiere las calamidades que trajo: tres meses de lluvia, trueno y relámpagos durante las cuales murieron hombres y ganado, el grano se secó en los campos y una peste siguió a un eclipse de sol y de luna.

suyo). Tanto si procedió de la manera descrita como si, simplemente, se limitó a recordar con vivacidad la aparición, su cometa naturalista armoniza perfectamente con la estética naturalista que caracteriza a la *Adoración* y al ciclo entero de los frescos.

¿Qué pudo haber llevado a Giotto a representar la estrella de Belén como un cometa? Nadie sabe, desde luego, si la estrella bíblica fue un cometa, una *nova*, una conjunción de planetas o si es apócrifa. Mateo es el único evangelista que menciona una estrella que llevó a los magos al lugar del nacimiento y no ofrece detalles: "Pues hemos visto su estrella en Oriente y hemos venido a adorarle... Y he aquí que la estrella que vieron en Oriente avanzó delante de ellos hasta que se detuvo al llegar al lugar en el que se encontraba el Niño. Cuando vieron la estrella, tuvieron una gran alegría" (Mat. 2, 2; 2, 9-10). La estrella aparece también mencionada en el apócrifo *Protoevangelio de Santiago,* pero tampoco aquí se encuentra una caracterización. Casi siempre se la presenta como una estrella pequeña, estilizada y totalmente imaginaria, frecuentemente con rayos de luz cuyo brillo descendía sobre el Niño Jesús, con lo que quería significarse que Dios bendecía el nacimiento.

Los pintores del Renacimiento dependían de reglas iconográficas bien establecidas. El estilo y la técnica podían variar, pero la elección de lo que podía representarse en una pintura, incluso el lugar que correspondía a cada figura y la disposición general del cuadro, dependía ampliamente de la tradición (o de la interpretación innovadora de la tradición) que, a su vez, daba forma a lo que el público esperaba. Giotto, por consiguiente, no hubiera sustituido fácilmente la estrella convencional por un cometa debido a un simple impulso, incluso como consecuencia de una vívida observación personal.

Existía, desde luego, una larga tradición literaria relativa a los cometas. Mucho antes de Giotto, escritores como Aristóteles, Virgilio, Séneca y el poeta romano Lucano, entre otros, habían especulado acerca de las apariciones de los cometas. Estos, en general, eran considerados con aprensión e interpretados como anuncios de catástrofes, epidemias o de la muerte de reyes. La interpretación variaba con las circunstancias. La aparición del cometa de Halley en 1066 pudo haber sido considerada por los ingleses como un indicio de la victoria de su propio rey, Harold, sobre Harold III de Noruega a fines de septiembre del mismo año, o bien como señal de su propia derrota en Hastings, al mes siguiente,frente a Guillermo de Normandía; los normandos, vencedores eventuales en el conflicto tripartito, consideraron, desde luego, el cometa como un anuncio del triunfo de Guillermo. Se creía que los cometas, cuando eran considerados augurios positivos, tenían un origen divino, y no natural: creados por Dios para una finalidad benevolente específica. Este punto de vista fue ilustrado por Egidio de Lessines en su tratado *Sobre la esencia, movimiento y significado de los cometas* (1264).

En la época en que se escribieron los Evangelios, a fines del siglo I, era corriente asociar la aparición de una "nueva estrella" con el nacimiento de un rey. Mucho más tarde, los escritos de los Padres de la Iglesia, que podían haber sido familiares a los contemporáneos de Giotto, ampliaron el relato de Mateo acerca de la estrella de Belén. Dos pensadores cristianos innovadores, Orígenes en el siglo III y Juan Damasceno en los siglos VII y VIII, observaron que la estrella descrita por el Evangelista tenía una naturaleza semejante a la de un cometa.

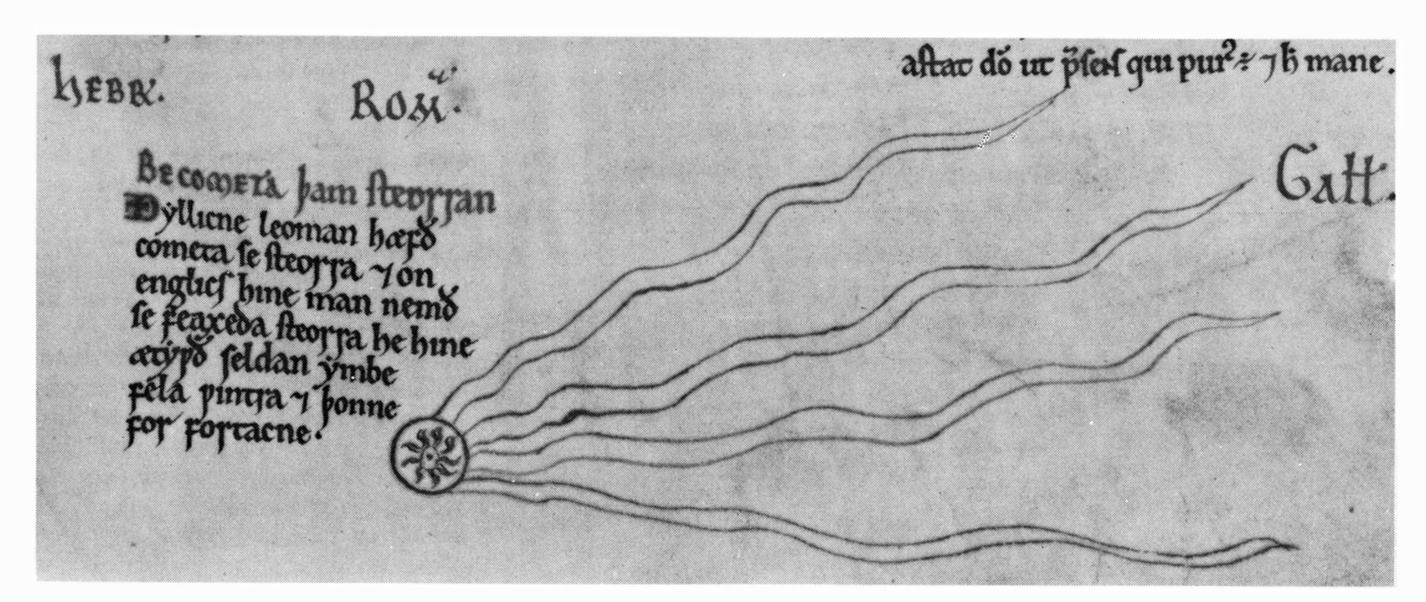
Orígenes escribió en su tratado *Contra Celso*: "Consideramos que la estrella que fue vista en Oriente fue una nueva estrella... como los cometas o esos meteoros [cometas y meteoros no se diferenciaban entonces] que se asemejan a vigas de madera, barbas o jarras de vino". Señaló que "hemos leído en el *Tratado sobre los cometas* de Jaeremon el Estoico que también en algunas ocasiones, cuando *algo bueno* va a suceder, se produce la aparición de cometas... Por consiguiente si, al comienzo de nuevas dinastías... surge un cometa..., ¿por qué habría de extrañarnos que en el momento del nacimiento de Aquel que iba a introducir una nueva doctrina para la raza humana... apareciera un cometa?". Orígenes reconoció que no existía ninguna profecía específica según la cual un cometa particular "aparecería en conexión con un reino o un tiempo concreto, pero en relación a la aparición de una estrella cuando se produjo el nacimiento de Jesús existe una profecía de Balaam recogida por Moisés a este respecto: "Aparecerá una estrella de Jacob a este efecto y un hombre surgirá de entre Israel (*Números* 24, 17)". (Es interesante observar que *The New English Bible* traduce así el pasaje: "Una estrella surgirá de Jacob, un cometa se elevará de Israel".)

Orígenes, por consiguiente, contribuyó con tres conceptos importantes a la leyenda de la estrella de Belén. Al repetir la profecía del Antiguo Testamento, indicó una conexión con la vieja tradición según la cual el nacimiento de profetas como Abraham y Moisés era

anunciado por una estrella, con lo que se acentuaba la continuidad entre los dos testamentos. Asoció claramente la estrella con un cometa. Por último estableció con claridad que los cometas podían tener carácter benéfico.

Juan Damasceno, un teólogo bizantino cuyos escritos fueron traducidos al latín en el siglo XII y que fue uno de los escasos padres griegos conocidos por los europeos de los siglos XII y XIII, relacionó la estrella con un cometa en su influyente *Exposición exacta de la fe ortodoxa*, aunque de manera menos directa: "Con frecuencia tiene lugar la aparición de cometas. Estos... no son ninguna de las estrellas que fueron creadas en el principio, sino que fueron formados en ese mismo tiempo por mandato divino y disueltos a continuación. Por tanto ni siquiera la estrella que vieron los Magos... es una de aquellas que [fueron] creadas en el principio. Esto resulta evidente ya que, a veces, su curso

EL TAPIZ DE BAYEUX, que conmemora los acontecimientos de 1066, registra la aparición del cometa de Halley en la primavera de aquel año. A la izquierda, una multitud de ingleses señalan el cometa estilizado situado en la parte superior. En una leyenda superior se lee: "Tienen miedo de la estrella". A la derecha, el rey Harold II, al ser informado del mal presagio, imagina fantasmagóricos barcos invasores que anuncian su derrota.



SALTERIO EADWINE, una colección manuscrita iluminada de los salmos. Fue copiado por el monje Eadwine del Salterio de Utrecht, más antiguo. A juzgar por las fechas en que vivió Eadwine, la prueba estilística y esta representación de un cometa en la parte inferior de una página, la copia se realizó en 1145 o poco después. La leyenda se refiere al brillo del cometa o "estrella con pelo", y observa que los cometas aparecen rara vez "y cuando lo hacen constituyen un portento". Por encima del cometa hay tres versiones latinas del Salmo 5: *Hebraicum*, *Romanum* y *Gallicanum*.



LA NATIVIDAD POR DUCCIO DI BUONINSEGNA es un pequeño entrepaño de la base de un gran retablo de altar, la *Maestà*, realizado entre 1308 y 1311 para la catedral de Siena; el entrepaño (con dos paños laterales que no aparecen aquí) se encuentra ahora en la National Gallery of Art en Washington. La estrella de Belén pintada por Duccio, en contraste con el cometa de la *Adoración* de Giotto, es una pequeña estrella con débiles rayos que se dirigen hacia el Niño. Duccio era mayor y rival de Giotto.



LA NATIVIDAD POR UN SEGUIDOR DE GIOTTO fue pintada hacia 1316 en la iglesia menor de San Francisco en Asís; el fresco, del que tenemos un detalle en esta ilustración, es una mezcla de motivos tomados de escenas de la Capilla de la Arena. Hay un cometa, pero es pequeño y estilizado, un pálido eco de Giotto, y ha sido subordinado a los rayos celestiales que descienden hacia el Niño.

se producía de este a oeste y, otras veces, de norte a sur. En un momento estaba oculta y, en otro, se manifestaba claramente: esto resulta poco armónico con el orden y la naturaleza de las estrellas". Esta descripción connotaba, de modo evidente, un cometa para el lector.

Los escritos de Flavio Josefo, un historiador judío del siglo I de nuestra era, apoyan la asociación de la estrella con un cometa. En *La Guerra Judía,* Josefo mencionaba conjuntamente el portento de una estrella con forma de ancha espada sobre Jerusalén y "un cometa que permaneció durante un año entero". Esto es, probablemente, una referencia confusa a la aparición del cometa de Halley en el año 66 de J. C., poco antes del comienzo de la revuelta judía contra Roma. Josefo anotó otras maravillas entre las que se encuentra una vaca que parió un cordero. Algunos autores cristianos leyeron más tarde estos detalles y los interpretaron como una alegoría de la estrella de Belén que anunciaba el nacimiento de Cristo (el cordero), mientras que los escritores judíos consideraron el pasaje como un presagio de la quema del Templo en el año 70 de J. C. (la aparición del año 66 pudo haber influido en Mateo, cuyo Evangelio fue escrito después de la caída de Jerusalén).

Existía una tradición muy popular en tiempo de Giotto que relacionaba la estrella bíblica con un cometa. En la *Leyenda Aurea* de 1275, obra ampliamente difundida (de la que se ha sugerido que podía ser la fuente de otros detalles del ciclo de la Capilla de la Arena pero que, hasta ahora, no se ha relacionado con este tema concreto), el teólogo y cronista genovés Jacobo de Vorágine describió la estrella de un modo que debió sugerir un cometa a sus lectores: "Era una estrella nuevamente creada y hecha por Dios... Fulgencio [un defensor de la ortodoxia en el siglo VI] dice: Se diferenciaba de las demás estrellas en tres cosas. En primer lugar en su situación, ya que no estaba fija en el firmamento sino que se encontraba suspendida en el aire cerca de la tierra. En segundo lugar en luminosidad... En tercer lugar en movimiento, ya que iba siempre delante de los reyes". Jacobo comentó también la etimología de "epifanía" y afirmó que derivaba de *epi* ("arriba") y *fanes* ("aparición") y, por consiguiente, identificó la Epifanía (adoración de Cristo por los Magos) con la aparición de una estrella divina. La palabra "Magos", por otra parte, deriva del griego *magoi* que, a su vez, procede de una palabra persa que denota a los brujos y astrólogos, a los sabios que estudiaban las estrellas.

Tanto la lengua como la literatura

apoyaban, por tanto, la existencia de una estrecha relación entre la estrella bíblica y la Adoración de los Magos, por una parte, y la aparición irregular y espectacular, por otra, de una estrella "nueva" y milagrosa que, muy probablemente, era un cometa. Conviene subrayar, también, que en tiempo de Giotto surgió una actitud más desapasionada —que podría calificarse casi de científica— en relación con los cometas. Un cronista contemporáneo con la aparición de 1301 concluyó su descripción con una afirmación de la libertad de la voluntad humana frente al poder de las estrellas. Padua era un centro en el que se cultivaban las matemáticas, la disciplina que, eventualmente, liberaría a la astronomía de las supersticiones de la astrología; ya a fines del siglo XIII los hombres de ciencia habían empezado a observar sistemáticamente el firmamento. (Galileo sería el titular de una cátedra de matemáticas en la Universidad de Padua en el siglo XVI.) Giotto, por consiguiente, trabajaba en el seno de una comunidad filosóficamente progresista que podía haber fomentado el espíritu de individualismo y objetividad con el que el pintor contempló el cometa y lo representó en su fresco.

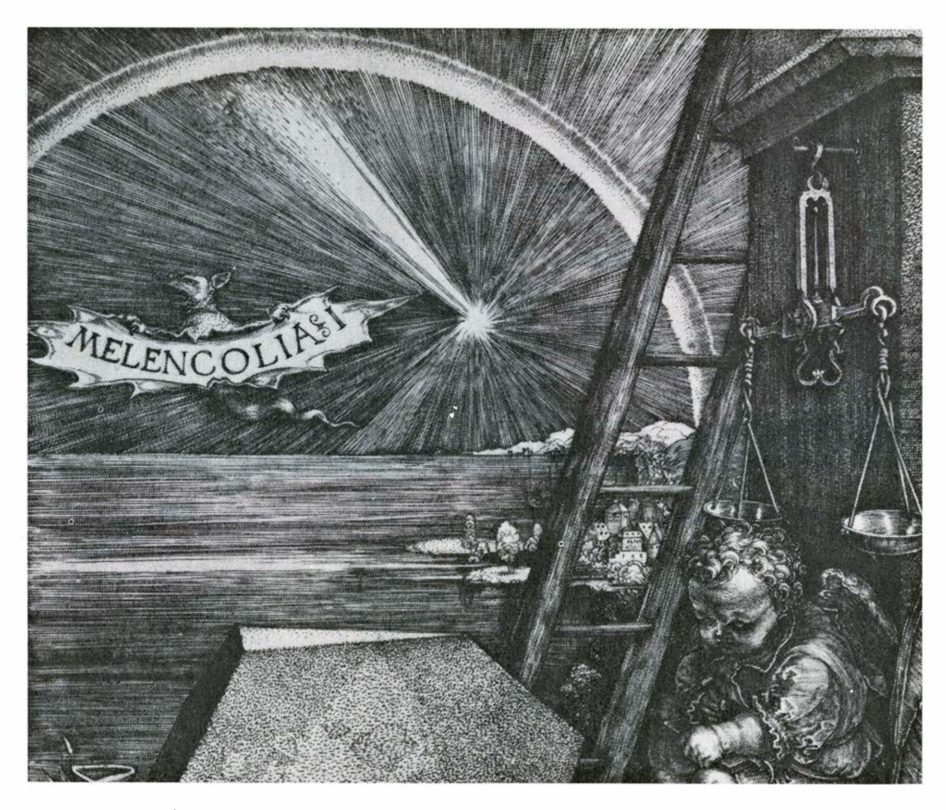
La naturaleza revolucionaria de la representación de Giotto queda de manifiesto si comparamos su cometa con otras ilustraciones de la estrella de Belén realizadas por sus contemporáneos y sucesores. El principal rival de Giotto en su tiempo, el pintor innovador, pero más bizantino, Duccio di Buoninsegna de Siena, realizó un inmenso retablo de altar, la *Maestà*, para la catedral de Siena entre 1308 y 1311. En la *Natividad* de Duccio, un entrepaño situado en la base del retablo, pintó una estrella de ocho puntas, estática y simbólica, conforme con la tradición iconográfica convencional, con finos rayos dorados que descendían hacia el Niño.

Hacia 1316, diez años después de la terminación del ciclo de la Capilla de la Arena, un seguidor de Giotto pintó un fresco de la Natividad en la iglesia menor de San Francisco en Asís, en el que combinaba en una sola escena una serie de motivos del ciclo de la Arena de Giotto. El artista anónimo sufrió una confusión en sus imágenes estelares y pintó un pequeño cometa convencional y un conjunto, separado y más prominente, de rayos dorados que descendían del cielo. (Pintó una estrella-cometa idéntica en la escena de la Adoración del ciclo, pero está mal conservada y apenas se ve.)

En contraste con el retrato analítico del cometa de Halley realizado por



ESCENAS DE LA INFANCIA DE CRISTO aparecen representadas en una página de unas concordancias iluminadas de los cuatro evangelistas realizadas en 1399 por Jacopo Gradenigo de Padua. La pintura contiene dos representaciones de la estrella de Belén. Sobre el pesebre (*arriba a la izquierda*) se encuentra una estrella estacionaria; por encima de los Magos (*arriba en el centro*) se muestra un cometa esquemático, en parte ángel y en parte estrella, con rayos que se proyectan hacia adelante y hacia atrás para formar la cola. Gradenigo no era un innovador: la naturaleza conservadora de su arte refleja su adhesión al principio de la "narración continua", con la presencia de escenas no contemporáneas.



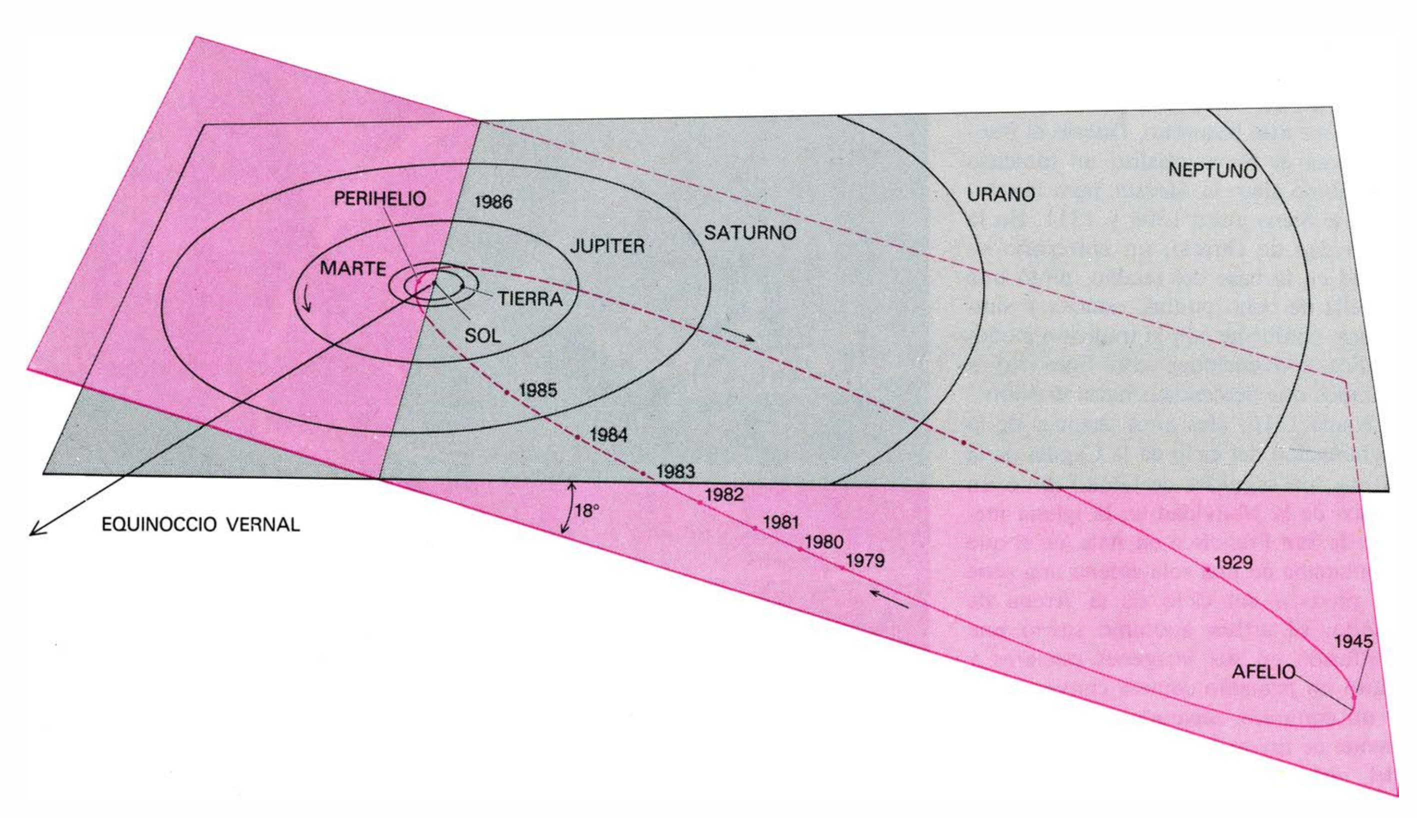
ALBRECHT DÜRER incluyó un cometa luminoso y dinámico, mucho menos naturalista que el de Giotto, en su famoso grabado *Melencolia I* de 1514. La ilustración es un detalle de la esquina superior izquierda del complejo conjunto, que se interpreta como un "autorretrato espiritual", que representa la melancolía del artista dada su incapacidad de unir teoría y práctica. El cometa y el arco iris pueden simbolizar la enfermedad, locura y acontecimientos milagrosos que se asociaban con la melancolía.

Giotto, el pequeño cometa de Asís es una estrella de ocho puntas estilizada, como la de Duccio, pero con una cola lineal en forma de mechón; parece haber sido incluido como una tímida imitación, basada en un vago recuerdo, del cometa de la Arena. O bien el artista no había estudiado nunca la aparición de un cometa o no tenía ningún deseo de representarlo de un modo convincente; sólo quería mostrar la estrella como un símbolo, basándose en su conocimiento de que Giotto lo había pintado como un cometa. Tal vez los rayos separados del cielo implicaban una crítica teológica a la representación poco ortodoxa de Giotto; al acentuar los rayos glorificantes, la intención del artista de Asís parece haber sido reforzar, de manera dogmática, la naturaleza divina del acontecimiento y la bendición que de él se desprendía. Tal vez la interpretación de Asís, menos atrevida, había sido requerida por la naturaleza muy pública de la importante peregrinación a esta iglesia. La Capilla de la Arena, por su parte, era privada, por lo que proporcionaba a Giotto un ambiente más tranquilo para un experimento que era tanto teológico como artístico.

El carácter avanzado del retrato que Giotto hizo del cometa se ve acentuado por la comparación con una obra muy posterior debida a otro artista de Padua. En 1399 Jacopo Gradenigo llevó a cabo un manuscrito iluminado, unas concordancias de los Evangelios. La estrella de Belén sobre los Magos, en su ciclo de la infancia, no es un auténtico cometa sino una representación simbólica, en parte ángel y en parte estrella, que culmina en una cola larga y majestuosa. Resulta muy esquemática y tiene poco que ver con la concepción naturalista de Giotto, aunque sea posterior en casi un siglo al fresco de la Capilla de la Arena. Los manuscritos iluminados tienden a reflejar las innovaciones anteriores realizadas en la pintura de carácter más monumental. Gradenigo no era un innovador: la naturaleza conservadora de su arte resulta evidente en su adhesión al viejo principio de la "narración continua" con la presencia, en la misma ilustración, de seis escenas temporalmente distintas. La representación, realizada por Gradenigo, de la estrella como un cometa, por estilizada que resulte, prueba que, a fines del siglo XIV, la interpretación de la estrella de Belén como un cometa había ganado aceptación en Padua.

A pesar de todo, el cometa no suplantó la interpretación más usual de la estrella en el arte europeo de manera general. Ni siquiera sabemos si Giotto la repitió ya que la pintura de la Capilla de la Arena es la única representación conservada de la estrella de Belén que puede, con seguridad, atribuírsele. Resulta un tanto fútil el especular sobre si la idea de representar la estrella de Belén como un cometa fue concebida por Giotto, por su patrono Scrovegni o por un teólogo asesor. Además de la cuestión de la innovación iconográfica, la importancia del cometa de la *Adoración* radica en su sorprendente naturalismo que incluso llega a superar al cometa representado por Dürer en 1514 en su *Melencolia I.*

Queda claro que en tiempo de Giotto un sector importante de la tradición asociaba la estrella de Belén con un cometa. Giotto se apoyó en esta tradición y en su propia observación del cometa de Halley en 1301, cuando pintó la *Adoración.* Al hacerlo, representó un cuerpo celeste que reflejaba una verdad naturalista y elevaba al mismo tiempo la intensidad emocional. Además, al pintar un cometa histórico, Giotto incrementaba el impacto contemporáneo de su representación de la Adoración. Fomentaba el que sus espectadores se identificaran con los testigos bíblicos del milagroso acontecimiento del nacimiento de Cristo: también ellos habían tenido la experiencia de la aparición espectacular de un cometa en 1301.



ORBITA DEL COMETA DE HALLEY (*en color*). **La órbita forma una elipse alargada con un eje mayor de unas 35,6 unidades astronómicas (1 U.A. es la distancia media entre la tierra y el sol) y un eje menor de unas 9 U.A. El movimiento del cometa es retrógrado, o sea que tiene lugar en dirección opuesta al movimiento de los planetas en torno al sol; el plano de la órbita está inclinado 18 grados con respecto al plano de la órbita terrestre (*gris*). El cometa alcanzará su perihelio, a una distancia de 6 U.A. del sol, el 9 de febrero de 1986. Los puntos coloreados muestran la posición en el perihelio y en la misma fecha en otros años; el punto negro muestra dónde se encontrará la tierra cuando el cometa alcance su perihelio.**

Juegos matemáticos

De cómo poseer psiquismo, incluso siendo caballo o cualquier otro animal irracional

Martin Gardner

Las artes de magia e ilusionismo tienen la finalidad de entretener a un auditorio realizando cosas que violan las leyes de la naturaleza, al menos aparentemente. Una mujer levita. Un elefante se esfuma. De la nada se materializan palomas inquietas. Las cucharas se doblan al tocarlas con el dedo. Y así sucesivamente.

Una pequeña rama de la magia, muy en el candelero estos días en razón del interés del público por los fenómenos psíquicos y las personas que los manifiestan –"psíquicos", como gustan de llamarse a sí mismos– se propone realizar cosas no imposibles, pero sí extremadamente improbables. Tales proezas requieren elementos que introduzcan azar, como barajas de cartas, dados, etcétera, por lo que existen considerables puntos de contacto entre los métodos de los charlatanes psíquicos y los métodos de los fulleros y los charlatanes honrados (magos).

La descripción completa de las relaciones entre magia y probabilidad, deteniéndonos a examinar los métodos imaginados para controlar naipes, dados, ruletas, monedas, bingos y ruedas de tómbola, ocuparía tantos volúmenes como una enciclopedia. Hay en todas estas variantes estratagemas y trampas de increíble ingenio. Por ejemplo, las ruedas de la fortuna de las barracas de feria, que giran en un plano vertical, pueden sesgarse con ayuda de un tablón flojo de la tarima. Un extremo de esa tabla hace palanca contra la punta de otra tabla vertical que pasa por detrás de la rueda, tabla en la que está perforado un agujero, atravesado por el eje de la rueda. El encargado de la tómbola se sitúa de pie, en el otro extremo de la barraca. Desplazando imperceptiblemente su peso de una pierna a la otra, puede actuar sobre el tablón flojo del piso, usándolo de palanca y empujar hacia arriba la tabla vertical, frenando de esta forma el eje de la rueda. Lo más hermoso de esta granujada es que ni siquiera desmontando por completo la barraca se puede demostrar que ha sido cometida deliberadamente.

Este mes seleccionaremos unos cuantos entre los centenares de procedimientos de que se valen los falsos psíquicos para potenciar su probabilidad al trabajar con naipes, sean éstos ordinarios o especiales para pruebas de percepción extrasensorial (PES). En la actualidad, ningún especialista o aficionado serio a la parapsicología permitiría que el sujeto pueda ver o tocar las cartas, pero en los comienzos, cuando se obtuvieron los resultados más sensacionales, era frecuente descuidar tales controles. Incluso hoy en día, en pruebas informales y en exhibiciones públicas por psíquicos desaprensivos, las cartas que servirán para la demostración de PES se encuentran totalmente a la vista.

¿Cómo podría apañárselas un psíquico espabilado para que su puntuación supere notablemente la que obtendría por puro azar? Un método consiste en marcar secretamente algunas cartas mientras opera con ellas o durante las pruebas preliminares. Puede hacerse de muchísimas formas. Los jugadores tramposos suelen recurrir al llamado "pringue" en los medios de la profesión, sustancia cerosa fácil de llevar en el dorso de un botón de la chaqueta, donde siempre estará al alcance de los dedos. Deja en el borde del naipe un tizne extraordinariamente débil e indistinguible de la suciedad que las cartas van adquiriendo con el uso ordinario. Otro método es hacer con la uña una diminuta marca en el filo de la carta. Un tercer procedimiento, rascar con la uña del pulgar unos dos o tres milímetros de borde, lo que produce una línea de color algo más claro. Es suficiente marcar unas cuantas cartas para alzar la puntuación hasta niveles significativos.

A veces, ni siquiera es necesario marcar los naipes para distinguirlos por el dorso. Cuando los naipes, sean ordinarios o de PES, se usan unas cuantas veces es fácil que adquieran toda suerte de minúsculas imperfecciones: restos de suciedad, ángulos ligeramente torcidos, etcétera. Por este motivo, en las partidas entre profesionales los naipes usados en cada juego se retiran sustituyéndolos por barajas nuevas. Supongamos que un jugador de póker tenga vista de águila y observe que el rey de corazones tiene una motita de suciedad en una esquina. Si en una partida de stud póker (que se juega con una carta cubierta y cuatro a la vista) reconoce la mota en la carta oculta de un contrario, ¿podría decirse que hace trampa?

No es de dominio público, ni siquiera entre magos y aficionados a la magia, que las cartas oficiales de PES actualmente utilizadas (con autorización de J. B. Rhine) son de "sentido único". Esto significa que examinando atentamente su dorso se observa que al girarlas 180 grados no muestran exactamente la misma figura. Por ejemplo, en el ángulo superior derecho de los naipes para PES puede haber o no haber una estrellita. Un psíquico podría apañárselas de muchas maneras para colocar en dirección única las 25 cartas del mazo oficial de naipes de PES, por ejemplo durante las pasadas de prueba. Así, podría intentar adivinar solamente las cartas que estén vueltas en un sentido, no intentándolo con ninguna de las restantes, que declarará no conocer. Estas cartas, de las que no hace ninguna conjetura, van siendo apiladas en un montón aparte. Tras la primera pasada (donde razonablemente puntuará más o menos al nivel de puro azar) antes de cerrar la baraja se le da vuelta a uno de los mazos. Todas las cartas están ahora en la misma dirección, y el psíquico preparado para obrar milagros.

La literatura sobre magia con naipes está llena de hábiles trucos basados en el principio de dirección única. El psíquico puede extender las cartas sobre la mesa, volverse de espaldas y pedirle a alguien situado hacia el extremo izquierdo de la mesa que seleccione una carta. Se le indica entonces a esa persona que le pase la carta a otra situada hacia el extremo derecho para que sea testigo del símbolo que contiene. La segunda persona devuelve la carta a la baraja extendida, recoge el mazo y lo baraja. Con las maniobras anteriores casi siempre se invierte el sentido del naipe. (Hay otros muchos procedimientos, algunos muy sutiles, para conseguir que una o más cartas queden en sentido contrario a las demás en una baraja inicialmente dispuesta en dirección única.) Las cartas se van colocando después en una larga fila. El psíquico se vuelve, y pasando lentamente la mano sobre la fila para "captar los fluidos", localiza fácilmente la carta elegida. Entonces la carta se vuelve boca arriba alzándola sobre su pie; en cambio, para retornarla a la baraja, se vuelve boca abajo alzándola sobre un

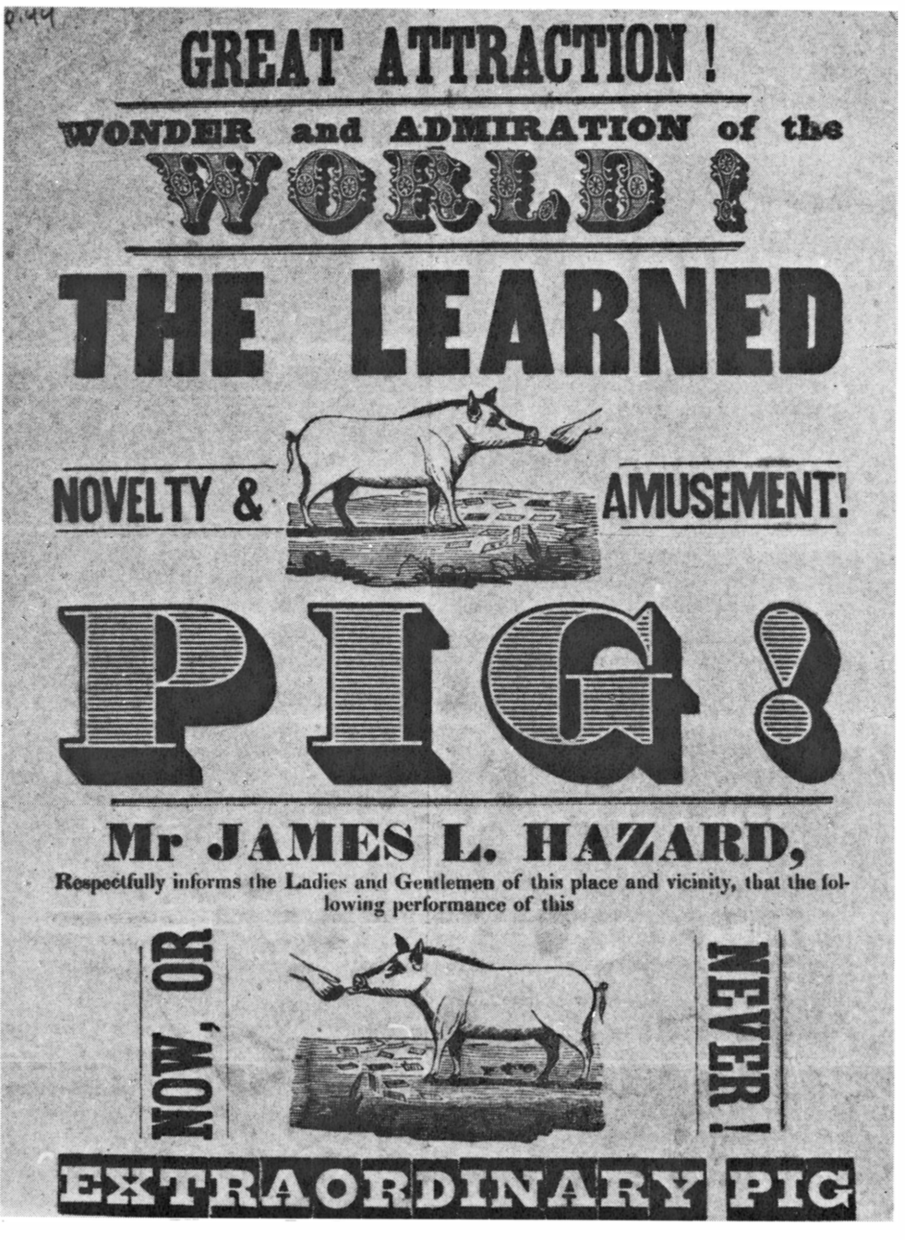
costado, con lo que todo el mazo vuelve a quedar en dirección única.

Para realizar parecidas proezas no es necesario tener toda la baraja de PES orientada en dirección única. Supongamos que los dorsos de las cartas del mazo estén en direcciones aleatorias, y que Vd. sea el psíquico. Se tienen cinco cartas sacadas al azar de la baraja y dispuestas en fila. Es suficiente memorizar en forma de número binario la combinación que forman, por ejemplo, 11010. Se sigue la estrategia recién descrita para seleccionar una carta y devolverla a la baraja, mientras Vd. permanece de espaldas. La carta es fácil de encontrar, y, memorizando las nuevas configuraciones, es posible repetir la exhibición tantas veces como se quiera.

Un mazo típico de cartas para experimentos de PES consta de 25 naipes; cinco para cada uno de los cinco siguientes símbolos: una estrella, una cruz, un círculo, líneas ondulantes, y un cuadrado. Es suficiente colocar 10 cartas correspondientes a dos símbolos en un sentido, y las 15 correspondientes a los tres símbolos restantes en el otro para poder conseguir puntuaciones impresionantes en la adivinación de los símbolos de las cartas que otra persona va sacando boca abajo de un mazo bien barajado. Para una baraja no preparada, la puntuación esperada es de cinco aciertos, suponiendo que no haya información realimentada, relativa a las cartas acertadas o falladas, hasta que se termina de dar todo el mazo. Colocando dos símbolos en un sentido, hay una probabilidad 1/2 de elegir correctamente uno de los dos, lo que da una esperanza de 10/2, o sea, cinco aciertos. Por otra parte, se tiene una probabilidad 1/3 de acertar correctamente en los otros tres símbolos, lo que da una esperanza de 15/3, o sea, de otros cinco aciertos. Dándole al mazo cuatro repasos completos, como suele hacerse en las pruebas de este tipo, es razonable esperar unos 40 aciertos en 100 intentos. Los especialistas en parapsicología consideran que una puntuación de 30 es señal de PES excelente, así que una cifra del orden de 40 es sencillamente sensacional. Es mucho más perfecta que el acierto total, que provocaría fortísimas sospechas de fraude. Vale la pena notar que si se marcase solamente uno de los símbolos, la esperanza para una prueba de 100 ensayos seguiría siendo de 40 aciertos. Sin embargo, para ello es necesario acertar siempre correctamente uno de los símbolos, lo que al quedar consignado en las actas de aciertos y fallos echaría a perder todo el juego.

Cuando el psíquico no puede ver las cartas en el desarrollo de la prueba, la forma más común de hacer trampa consiste en servirse de lo que los magos llaman "un soplón", o sea, una persona de confianza que puede observar disimuladamente lo que ocurre y que envía discretamente señales al psíquico mediante cualquiera de las infinitas –si bien no muy difundidas– técnicas de comunicación. (Uno de los procedimientos más difíciles de descubrir consiste en enviar impulsos radioeléctricos que el psíquico recibe a través de un diminuto dispositivo electrónico inserto en el ano. Pocos exámenes de presuntos psíquicos han tenido en cuenta esta posibilidad.)

El empleo de soplones ocultos ha sido corriente en la historia de los fenómenos psíquicos. Casi todos los animales capaces de "leer el pensamiento", que en el pasado contaron con caballos, perros e incluso cerdos, y que protagonizaron famosos números de vaudeville, se basan en señales sonoras secretas que estas especies de grandes orejas y sentido auditivo muy superior al humano pueden captar sin dificultad. Por ejemplo, el jugador ha reconocido por cualquiera de los procedimientos ordinarios de ilusionismo la carta seleccionada. El animal va entonces "revisando" una por una las 52 cartas de una baraja extendida en el suelo o expuesta sobre un gran caballete; cuando llega a la carta correcta, el animal la señala con el hocico, con la pata o la recoge con los dientes. Como es obvio, el animal recibe la indicación de la carta oportuna mediante una señal que solamente él puede detectar, y que puede ser casi cualquier cosa, desde un



Anuncio publicitario de un cerdo capaz de leer el pensamiento. (Colección de Milbourne Christopher)

suave sorbido de nariz hasta el chasquido de las uñas de una mano, que puede estar metida en un bolsillo o doblada a la espalda. Tampoco es necesario que sea una persona del escenario quien dé la señal; puede muy bien ser cualquier "espectador" de primera fila.

Si el animal está convenientemente amaestrado, es capaz de ir recogiendo letras del abecedario y componer una palabra, o de seleccionar tarjetas numeradas que den la solución de algún problema aritmético propuesto por un espectador. Muchos libros antiguos describen con gran detalle cómo entrenar animales para ejecutar tales actos. Uno de los mejores es *Haney's Art of Training Animals* (Jesse Haney & Co., 1869), quien recomienda que las señales se hagan con las uñas. Otro de ellos es *The Expositor: Or Many Mysteries Unravelled,* por William Frederick Pinchbeck (publicado por cuenta del autor, 1805), que contiene una excelente sección dedicada al entrenamiento de cerdos por el método del sorbido.

En su juventud, J. B. Rhine fue totalmente burlado por un caballo de Virginia llamado Lady Wonder, cuyas facultades psíquicas, particularmente en el campo de lectura del pensamiento, fueron entusiásticamente descritas por Rhine y su esposa en *The Journal of Abnormal and Social Psychology* (vol. 23, n.º 4., págs. 449-466, enero-marzo, 1929). Más adelante, el mismo Rhine, cuyo espíritu crítico se había afilado en el ínterin, repitió sus pruebas con Lady Wonder, y descubrió que su entrenador le hacía señales. Hasta el día de hoy, Rhine insiste en que Lady Wonder tenía genuina percepción extrasensorial, que la perdió, y que entonces su entrenador empezó secretamente a hacerle señales. (Pueden consultarse los detalles en mi libro *Fads and Fallacies in the Name of Science,* publicado por Dover Paperbacks, págs. 351-352.)

Por lo general, la lectura de las actas oficiales de los experimentos de parapsicología no permiten saber si estaban presentes soplones o no. Repasando los artículos de Harold E. Puthoff y Russell Targ donde se refieren los tests de clarividencia a que fue sometido el mago israelí Uri Geller con ocasión de su visita al Stanford Research Institute, no se encontrará en ellos ninguna indicación de que el mejor amigo de Geller, llamado Shipi Strang, estuvo siempre presente durante el desarrollo de las pruebas. Cuando este hecho salió a relucir, Puthoff y Targ accedieron a confirmar que, efectivamente, Targ se encontraba allí, aunque haciendo constar que siempre estuvo cuidadosamente "alejado de la zona de objetivos". Con ello pretendían decir que Strang no se encontraba en la sala donde las imágenes objeto iban siendo seleccionadas de forma aleatoria para ser "enviadas" a Geller. Pero resulta que Strang, a quien apenas si se le prestó atención, dispuso de toda libertad para rondar por donde quiso, y hay muchas formas mediante las que pudo haberse enterado de los objetivos y haber enviado las correspondientes señales a Uri. El contraste entre las elípticas y asépticas descripciones de los experimentos, y las caóticas condiciones en que se desarrollaron, narradas por John Wilhelm en *The Search for Superman* (Pocket Books, 1976) y James Randi en *The Magic of Uri Geller* (Ballantine Books, 1976), es, cuando menos, sorprendente. Que Strang hizo no pocas veces de soplón para Uri Geller está bien probado, por las declaraciones de una hermana de Geller en una entrevista concedida a un periodista israelí, y también por el primer agente de Geller, Yasha Katz, que así lo manifestó el año pasado en una entrevista para la televisión italiana.

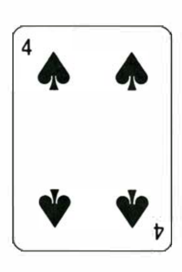
En el caso de que el psíquico se proponga adivinar mediante PES cartas que no puede observar directamente, el soplón puede transmitirle valiosa información aun sin necesidad de preparar la baraja, ni marcar las cartas u orientarlas todas en la misma dirección. Sólo es necesario que el soplón le vaya indicando el verdadero símbolo de la carta después de aventurarse la conjetura y descubrirse el naipe. Esta información le permite al psíquico puntuar muy por encima del puro azar. Un psíquico que dispusiera de información completa del verdadero valor de cada carta después de cada conjetura suya podría, si aplicase la estrategia óptima, aumentar su puntuación media para 25 cartas de cinco a 8,65. [Véase Ronald C. Read, "Cardguessing with Information - a Problem in Probability", *The American Mathematical Monthly,* vol. 69, n.º 6, págs. 506-511; junio-julio, 1962.]

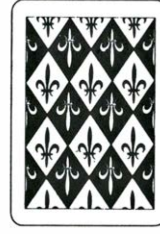
Si la realimentación solamente informase al psíquico de si ha tenido un acierto o un fallo, todavía podría poner en práctica una estrategia capaz de darle una puntuación media de 6,63 aciertos. [Véase Persi Diaconis, "Statistical Problems in ESP Research", *Science,* vol. 201, n.º 4351, págs. 131-136; 14 de julio de 1978.] En muchas pruebas clásicas de PES es imposible saber por el informe publicado si el sujeto dispuso de algún tipo de realimentación, o si se encontraban presentes amigos o personas que pudiesen enviarle información haciéndole señales secretamente.

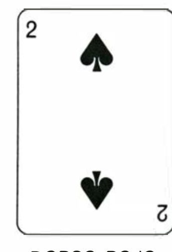
En casi todos los juegos de cartas, un jugador hábil puede extraer valiosas informaciones de los naipes ya jugados. Es evidente que así ocurre en el bridge y en el stud poker. En el blackjack (o las 21) que es un juego de casino, un jugador que consiga retener en la memoria las cartas repartidas puede realmente ganar de forma sistemática, apostando fuerte cuando la suerte le sonría, y reservándose cuando la tenga en contra. Los jugadores de este tipo son llamados "calculistas"; algunos de ellos han tenido tanto éxito que la mayoría de los casinos no les permiten entrar en partidas de este juego; lo que difícilmente puede estimarse justo, pues el jugador calculista no hace trampa en ninguno de los sentidos razonables del término.

Al valorar la realimentación producida por el conocimiento de las cartas repartidas la intuición puede descarriarse de forma increíble. El más sorprendente ejemplo de este tipo que yo conozco es un juego de apuestas, descubierto por Robert Connolly, matemático de la Universidad de Cornell, quien dio


CINCO DE CORAZONES

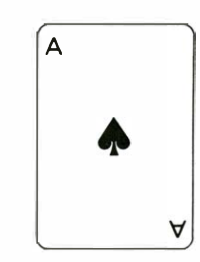



TRES DE ESPADAS



DORSO ROJO


SEIS DE ESPADAS

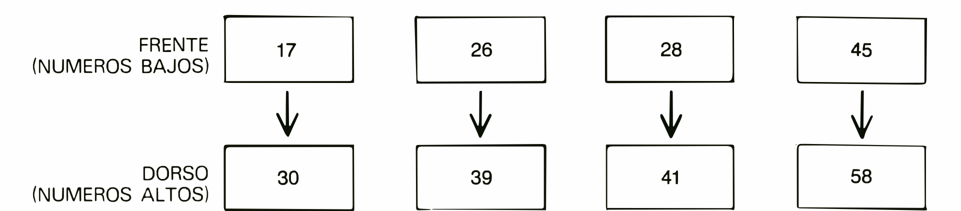


Truco de predicción de la carta roja

que hablar el año pasado, cuando demostró la falsedad de una conjetura muy famosa sobre poliedros. Supongamos que todas las caras de un poliedro sean rígidas, pero que estén articuladas por bisagras a lo largo de sus aristas. ¿Podrá "flexionarse" el poliedro? Es decir, ¿podrá cambiar de forma? Se sabía desde 1813 que si el poliedro es convexo la flexión es imposible, y se había conjeturado que lo mismo sucedería con todos los poliedros, siempre que su superficie no se corte a sí misma. Connolly encontró un contraejemplo formado por 18 caras triangulares. (Puede verse una descripción sin tecnicismos en su artículo "A Flexible Sphere", en *The Mathematical Intelligencer,* vol. 1, n.º 3, págs. 130-131; 1978.)

Connolly ha llamado a su juego "Diga Roja". El banquero baraja un mazo ordinario de 52 cartas (baraja francesa) y va sacándolas lentamente boca arriba. Las cartas ya extraídas quedan a la vista del jugador, quien puede examinarlas cuando lo desee. En un momento que el jugador decide libremente, puede decir "Roja". Si la próxima carta es roja, gana la partida y la cantidad apostada; de lo contrario pierde. Es obligatorio cantar "roja" antes de que termine el reparto, aunque se puede esperar hasta la última carta. Suponiendo que el jugador aplique la estrategia óptima de que disponga, aprovechando al máximo la realimentación producida por las cartas que van siendo extraídas, ¿qué ventaja debe darle el banquero para que el juego sea justo? El jugador tiene que anunciar cuánto va a apostar antes de comenzar el juego. Daré el mes que viene la sorprendente solución de Connolly.

Otro de los secretos mejor guardados de que se sirven los charlatanes psíquicos para incrementar sus posibilidades de éxito consiste en utilizar lo que Diaconis llama puntos finales múltiples en su artículo antes mencionado. Los ilusionistas los conocen con el nombre más informal de "salidas". En esencia, la idea consiste en no especificar por adelantado cuál es el objetivo exacto que se propone conseguir, dejando que el resultado dependa de lo que ocurra. Como dice Diaconis, la probabilidad de que se produzca algún tipo de coincidencia es mucho mayor que la probabilidad de que se produzca una coincidencia especificada por adelantado. Diaconis describe el trabajo de un psíquico conocido en la literatura sobre el tema por "B. D.", cuyas proezas con naipes de juego fueron favorablemente comentadas en el *Journal of Parapsychology* publicado por



Truco de predicción de sumas de Shigeo Futigawa

Rhine. Diaconis, que ha sido ilusionista –y muy hábil– con juegos de cartas, estuvo presente como observador durante una de las exhibiciones de B. D. en la Universidad de Harvard. Inmediatamente se dio cuenta de que el principal recurso de B. D. era su gran habilidad en las salidas.

Si bien aquí no puedo describir con detalle de qué forma hacen amplio uso de salidas los ilusionistas que pretenden poseer psiquismo, si puedo ilustrar la técnica con una anécdota más dos notables trucos de "precognición" que permitirán a los lectores asombrar a sus amigos. Paul Curry, ilusionista aficionado de New York, gusta de referir una ocasión en que le pidieron a otro mago amigo suyo que realizase algunos trucos de cartas en una fiesta. Su amigo observó que la baraja que le proporcionaron tenía una carta, el ocho de tréboles, pongamos por caso, malamente estropeada. Como esa carta entorpecería sus manipulaciones, subrepticiamente la retiró del mazo y se la guardó en el bolsillo. Durante su exhibición pidió al auditorio que dijese una carta. Una señora gritó: "El ocho de tréboles". Sin pensarlo dos veces, el mago le contestó: "Lo lamento señora, pero tendrá Vd. que pedir otra carta. Verá, es que antes de empezar me di cuenta de que el ocho de tréboles estaba muy arrugado y..." El mago se interrumpió entonces, y casi se desmaya al caer en la cuenta de que había perdido una formidable ocasión de obrar un milagro: hacer que la carta mencionada se esfumase del mazo y a continuación se materializase en el bolsillo, deteriorándose por el camino.

Quizá cueste creerlo, pero hay ilusionistas que tienen a punto trucos de cartas (con baraja arreglada, o sea ordenada convenientemente) con 52 salidas diferentes, de acuerdo con la carta que sea elegida. La carta pedida puede aparecer al comienzo o al final del mazo, o ser la única que está al revés; puede que un duplicado de la carta esté clavado al techo con una chincheta, o quizás al abrir el pañuelo de bolsillo aparezca una gran reproducción del naipe; tal vez se encuentre bajo el cojín del asiento de un espectador, o ser la única carta que falta en el mazo; o tal vez resulte que al lanzar la baraja completa contra la persiana al alzarla se encuentre la carta pegada al vidrio de la ventana, y así ad infinitum.

El siguiente truco, muy ingenioso, tiene seis naipes y seis salidas; hay numerosas variantes. La versión que ofrezco fue preparada por Tom Ramson, de Toronto. La ilustración de la página opuesta muestra cómo deben alinearse las seis cartas. Puede decírsele honradamente al auditorio que las seis cartas tienen valores de 1 a 6. Todas menos una tienen reverso azul. El dos de espadas tiene reverso rojo, y está boca arriba. Todas las cartas son del palo de espadas, excepto el cinco de corazones, que evidentemente estará boca abajo. Se escribe en un papel: "Vd. elegirá la carta roja", y se deja la predicción boca abajo.

Se solicita un número de 1 a 6. Si se desea, se le puede dar a un espectador un "dado invisible" y pedirle que haga como si fuera a lanzarlo, para ver qué número sale. He aquí las seis salidas correspondientes a cada uno de los seis números:

1. Cuente desde la izquierda hasta la primera carta. Vuélvala boca arriba, para que se vea que tiene figura roja. Déles también la vuelta a las cartas que aún estén boca abajo, para que todos puedan ver que tienen figuras negras.
2. Déle la vuelta al dos de espadas, para mostrar que su dorso es rojo. Vuelva las otras dos cartas que aún están boca arriba para mostrar que todas las demás cartas tienen dorso azul.
3. Pídale a un espectador que cuente hasta la tercera carta desde su izquierda (del espectador). Se termina como antes.
4. Cuente hasta la cuarta carta desde la izquierda de Vd. Termine como antes.
5. Vuelva todas las cartas dispuestas boca abajo para revelar que la carta de quinto lugar es la única que tiene figura roja.
6. Pídale al espectador que cuente desde su izquierda hasta la sexta carta. Termine como antes.

No importa el número que se elija, su predicción siempre será exacta. Evidentemente, el truco no puede repetirse con el mismo auditorio.

Actualmente se encuentra a la venta

en las tiendas japonesas otro ingenioso truco con varias salidas, inventado por un profesor de matemáticas aficionado al ilusionismo, Shigeo Futigawa, de Tokio.

Se necesitan cuatro tarjetas idénticas, en blanco por los dos lados. En una de ellas se anota el número 17, y en su dorso el número 30. En las otras se escriben los pares 26/39, 28/41 y 45/58. En la primera fila de la ilustración de la página anterior se han mostrado cuatro fichas como éstas, con los números pequeños hacia arriba. Es necesario aprenderse de memoria estos cuatro números de forma que sea posible identificarlos rápidamente, o si se prefiere, marcarlos con un puntito, para distinguirlos de los números más altos ahora situados en el dorso de las tarjetas.

Se le dan las cuatro fichas a un espectador, pidiéndole que las revuelva a conciencia, dándoles la vuelta tantas veces como quiera, y que coloque después las fichas sobre la mesa; de forma que los cuatro números visibles están tomados al azar. Antes de que el espectador lo haga, se escribe una predicción de la suma total, que se pone boca abajo en una esquina.

Cuando las fichas así barajadas se ponen en la mesa, hay solamente tres posibilidades. En cada una de ellas se sigue un procedimiento diferente:

1. Las cartas muestran dos números bajos y dos números altos. Así ocurre en tres de cada ocho ensayos, suponiendo que se repita suficientemente. Sume Vd. estos cuatro números. La suma será 142. Esa es la cifra escrita en el papel; haga que un espectador lo compruebe.

2. Las cuatro cartas pueden ser altas, o las cuatro bajas. Así sucede en una de cada ocho veces. Póngase en pie, vuélvase de espaldas, y pídale a alguien que las revuelva un poco más, volviendo cualesquiera dos de las tarjetas. Evidentemente, al hacerlo así quedan dos altas y dos bajas, y se termina como en el caso anterior.

3. Las fichas pueden mostrar bien tres altas y una baja, bien tres bajas y un alta. Quizá le resulte sorprendente, pero ésta es la situación más frecuente, y se presenta con probabilidad de exactamente 1/2. Cuando así suceda, se toma la carta singular (tanto si ésta es la única carta baja como si es la única alta), y conforme Vd. le da la vuelta se dirige al auditorio diciendo: "Fíjense por favor en que los números situados en los dos lados de la ficha no son iguales. Si hubiesen colocado la ficha del revés, la suma de los cuatro números visibles hubiera sido totalmente distinta." Entonces, poniéndose de espaldas, le pide a un espectador que le dé la vuelta a dos fichas cualesquiera para que el resultado sea todavía más azaroso. Como al hacerlo así quedan iguales números de cartas altas y bajas, se puede terminar como en el primer caso.

En breve, adoptando este procedimiento siempre resulta una suma de 142. Así que no se puede fallar. Obsérvese que la diferencia de cada par de números de cada tarjeta es siempre 13. Dejo al cuidado del lector el trabajo de demostrar algebraicamente que el truco siempre funcionará. La diferencia puede ser cualquier otro número. El total predicho será siempre la suma de los cuatro números pequeños más el doble de la diferencia. De esta forma es sencillo preparar otras tarjetas con números distintos de los ocho antes dados.

¿Es posible preparar cuatro fichas con ocho números de forma que el truco se realice exactamente como antes, pero multiplicando en vez de sumando? Es necesario saber predecir el resultado del mismo modo que antes se predijo la suma, siguiendo los mismos tres procedimientos de salida. La respuesta es afirmativa, y Futigawa ha diseñado el correspondiente juego de fichas, variante de su truco anterior. El mes que viene describiré este segundo conjunto, y explicaré el secreto de su construcción.

No están de más un par de observaciones sobre la sección del mes de marzo, dedicada a los sistemas de circunferencias tangentes. Benjamin L. Schwartz me comunica que en su artículo "Separating Points in a Square", en *Journal of Recreational Mathematics* (vol. 3, n.º 4, págs. 195-204; octubre de 1970) se da una demostración de la empaquetadura óptima de seis círculos unitarios en un cuadrado. El problema de empaquetamiento de circunferencias en la forma en que fue presentado es equivalente al problema de situar n puntos dentro de un cuadrado de forma que la distancia mínima entre pares de puntos sea lo mayor posible. Y mi descripción de la espiral de H. S. M. Coxeter, donde se decía que es única, induce a error, pues solamente es única cuando la razón entre circunferencias consecutivas se especifica de antemano. Pueden verse más detalles en el artículo de Coxeter mencionado en la bibliografía de este número. Por otra parte, el Dr. José de Prada ha enviado a *Investigación y Ciencia* parte de un interesante trabajo inédito, sobre familias de esferas tangentes, del que se deducen, como caso particular, algunas propiedades de las cadenas de Steiner y de los trenes de circunferencias en la arbelos.

Taller y laboratorio

Cómo medir el tamaño de la tierra con sólo una mira o un cronómetro

Jearl Walker

Por distintos conductos me han llegado dos ingeniosos procedimientos para medir el radio de la tierra. Los remiten Joseph L. Gerver, de Honolulú, y Denis Rawlins, de San Diego. Ambos procedimientos, además de ingeniosos, son lo suficientemente sencillos como para que cualquier aficionado pueda repetirlos. El lector se divertirá viendo cómo puede obtener valores muy cercanos al radio real de la tierra y cómo la introducción de sutiles variaciones en el método alteran los resultados.

El método de Gerver precisa de tres piedras, una cuerda y una regla. De acuerdo con sus indicaciones, se debe contar también con un cerro desde donde podamos divisar el horizonte del mar en dos direcciones opuestas. La altura ideal de esta colina o promontorio sería de 30 a 300 metros sobre el nivel del mar; desde ella se divisará claramente una zona de playa. Gerver hizo sus trabajos en Makapuu Point, en el extremo oriental de Oahu. El lector deberá tratar de encontrar una península similar. Puesto que parte del procedimiento implica la observación de dos piedras situadas en la playa, es necesario poder ver también la orilla. La última exigencia es una piedra puntiaguda o pivote que colocaremos como señal en la cima del cerro.

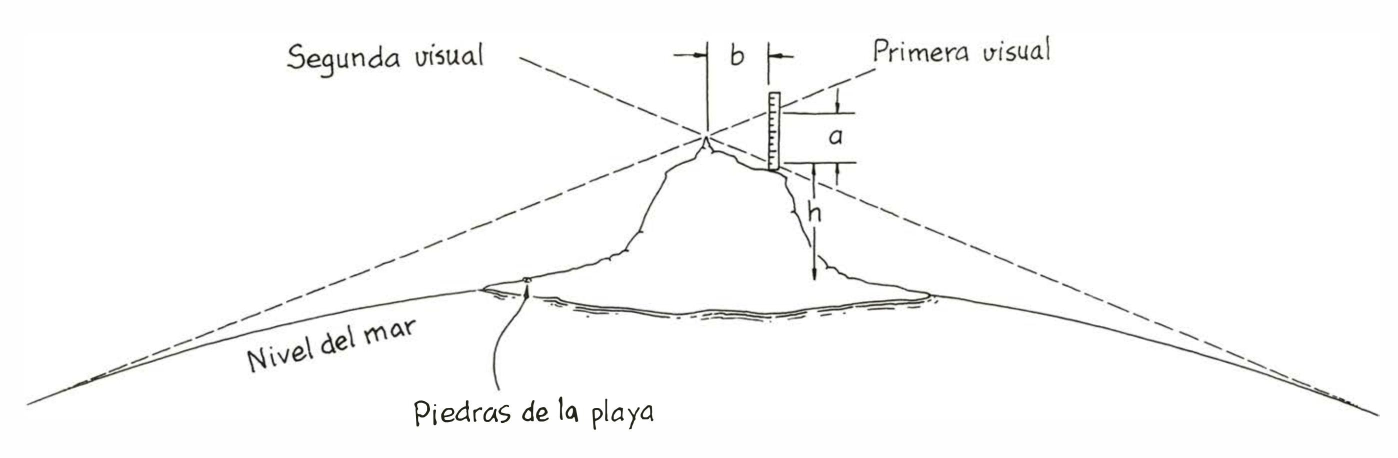
Situemos la regla en posición vertical a una distancia de unos 50 centímetros del pivote y midamos exactamente dicha distancia con la cinta. Gerver llama b al valor de la distancia. Para realizar la primera medición, el lector ha de colocarse de suerte que entre él y la piedra se encuentre la regla; irá ensayando la posición hasta que consiga ver alineados el horizonte, el extremo de la roca y un punto de la regla. Apuntará luego el punto de la regla en que la línea de visión corta la mira. Después debe moverse hacia la piedra, ajustando su línea de visión hasta que el extremo superior de la piedra esté alineado con el horizonte en la dirección opuesta. De nuevo ha de anotar el punto de intersección de su línea de visión con la mira (a a la distancia entre los puntos de intersección).

Por último hay que conocer la altura del observador sobre el nivel del mar. Para ello coloquemos dos piedras en la playa de manera que equidisten del observador y que, a su vez, estén separadas por una distancia conocida. Situados en el punto de observación y manteniendo la regla en el extremo de nuestro brazo y en posición horizontal, divisaremos las dos piedras, debiendo anotar la separación aparente que tienen sobre la mira. Giraremos luego la muñeca hasta colocar la regla en posición vertical para que, al divisar las piedras y la línea del horizonte, podamos anotar la distancia que separa las piedras del horizonte medida en unidades iguales a la distancia aparente entre las dos piedras. Multiplicando dicha distancia aparente [de las piedras al horizonte] por la separación real entre las dos piedras obtenemos la altura del observador sobre el nivel del mar, a la que llamaremos h. Finalmente calcularemos el radio de la tierra haciéndolo aproximadamente igual a $8h\,b^2/a^2$.

Por ejemplo, en una de las observaciones de Gerver, la distancia entre los dos puntos de intersección en la regla fue de aproximadamente 2,5 centímetros, la distancia entre la piedra y la mira 3,7 metros y el punto de observación estaba situado a 63,4 metros sobre el nivel del mar. Con estos datos el radio de la tierra que se calcula es de 10.500 kilómetros, que es apreciablemente mayor que el radio real de la tierra en el ecuador, que es de 6378 kilómetros (el radio polar real es de 6357 kilómetros). Repitiendo las observaciones se pueden reducir los errores aleatorios incorporados a las mediciones, pero los distintos errores sistemáticos permanecen.

¿Cómo funciona la fórmula de Gerver? La base geométrica, que se muestra en las ilustraciones de las dos páginas siguientes, implica cuatro triángulos fundamentales. Los cuatro son triángulos semejantes, esto es, tienen ángulos iguales. Gerver los empleará para establecer diferentes relaciones en las que están implicadas las tres medidas llevadas a cabo durante el experimento (a, h y b) y r, el radio de la tierra. Con ellas pudo despejar r.

Aquí vamos a esquematizar su demostración, pero dejaremos los detalles para el lector. A partir de los triángulos *1* y *2* podemos escribir la siguiente relación $CO/r = r/(r + h)$. A partir de los triángulos número *2* y *3* podemos decir que $AC/BC = CO/AC$. A partir de los triángulos *2* y *4* podemos establecer que $b/(a/2) = CO/AC$. Por último, tenemos que $BC = r + h - CO$. Eliminando CO, AC y BC a través de las correspondientes sustituciones en las relaciones anteriores podemos obtener r despejada. En un determinado momento del desa-



Método de Joseph L. Gerver para calcular el radio terrestre

rrollo de los cálculos, Gerver hace la aproximación de que la altura de la colina, en la que el observador se sitúa, es despreciable comparada con el diámetro terrestre. Dicha aproximación introduce un error en los cálculos de menos del 0,01 por ciento.

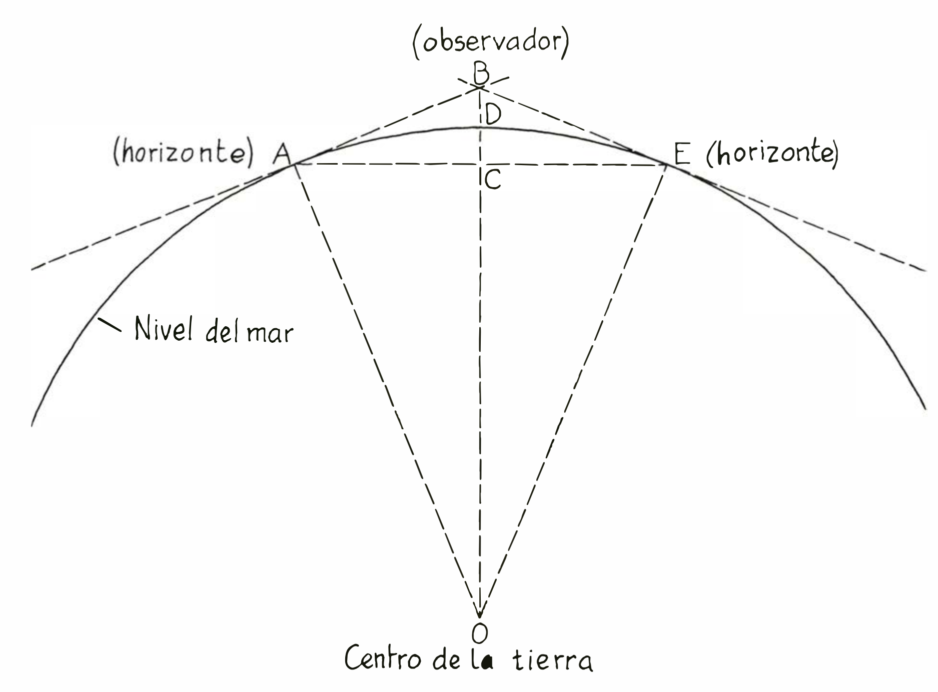
La altura h de la posición del observador sobre el nivel del mar se mide dividiendo los pivotes plantados en la playa. Gerver equipara la relación entre h y la separación real de las piedras a la relación existente entre la separación aparente de las piedras y la distancia aparente que hay entre las piedras y el horizonte oceánico más próximo a ellas. Como en el esquema de Gerver esta distancia se mide en unidades equivalentes a la separación aparente entre las mismas, se obtiene automáticamente la segunda de las dos relaciones propuestas. Cuando multiplicamos la segunda relación por la separación real entre las piedras, obtenemos la altura del observador sobre el nivel del mar.

Además de los errores fortuitos presentes en las observaciones, influyen también en el resultado varios errores sistemáticos. Uno de ellos provoca la curvatura de los rayos de luz tangentes al horizonte, que elevan ligeramente el campo de visión. Este efecto de refracción provocado por la atmósfera introduce un grado de error en los cálculos de Gerver demasiado elevado, aproximadamente del 20 por ciento. Hablaremos de este efecto más tarde, cuando describamos el método de Rawlins para medir el tamaño de la tierra.

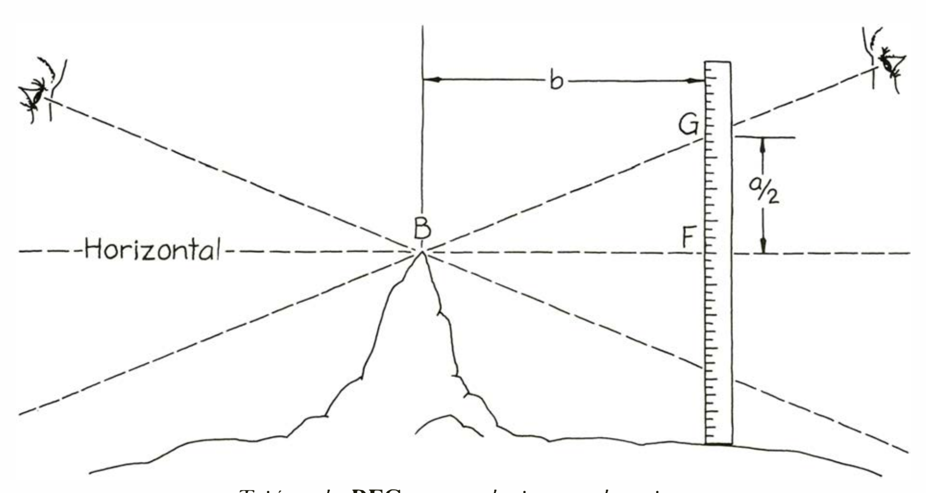
Otro factor a considerar son las olas del agua, que inciden al elevar el horizonte de nuestro campo de visión. Para eliminar ese efecto, se resta la mitad de la altura media de la ola (la mitad de la distancia entre cresta y valle dada por el Servicio de Guardacostas u otro organismo) del valor de h obtenido por el método de Gerver.

Otra causa de error aleatorio es la imposibilidad de ajustar perfectamente la alineación del horizonte del océano con el extremo superior de la piedra y con la marca en la regla. No se trata sólo del tipo de error que lleva consigo cualquier medición de una longitud, sino que además el error se acentúa porque esos elementos en particular no producen en la retina una imagen definida.

Y es fuente de error, asimismo, el cálculo de la altura de la atalaya de observación sobre el nivel del mar. El método de estimación de dicha altura, basado en la distancia existente entre las dos piedras situadas en la playa, supone que la línea que une el horizonte con las piedras de la playa es horizontal; la realidad es que está inclinada hacia abajo debido a la curvatura de la tierra. Dicho ángulo, que recibe el nombre de ángulo de inclinación, es proporcional a la raíz cuadrada de la altura del observador sobre el nivel del mar.



Geometría del método de Gerver



Triángulo **BFG** *entre el pivote y la mira*

Poniendo sumo cuidado en las observaciones, y multiplicándolas, el lector podrá aproximarse bastante al valor real del radio de la tierra, si sigue el método de Gerver. (Obviamente, no conseguirá medir el radio ecuatorial del planeta, sino el radio de curvatura local de la superficie del océano.)

Rawlins publicó su método para determinar el tamaño de la tierra en el número de febrero de la publicación *American Journal of Physics*. Básicamente, lo que Rawlins nos dice es que observemos dos puestas de sol en un corto intervalo de tiempo, una de ellas tendidos en la playa y otra poniéndonos de pie en el mismo sitio. Midiendo el tiempo transcurrido entre los dos ocasos se puede calcular el radio de la tierra con una aproximación del 10 por ciento más o menos.

Para poder llevar a la práctica el método de Rawlins, el observador deberá disponer de una clara visión del ocaso desde una playa donde se divise el mar en relativa calma o bien un gran lago. Para medir el tiempo conviene utilizar un cronómetro, pero puede servir también el segundero de un reloj de pulsera. Hay que prestar suma atención a la manera de mirar al sol. Si fijamos la mirada en él largo tiempo pueden resultar dañados los ojos. El mayor peligro viene de las radiaciones invisibles, que pueden perjudicar sin que provoquen molestias en el momento. Sería conveniente rehuir

la visión directa del disco solar hasta que esté en su mayor parte por debajo de la línea del horizonte. Debe seguirse la trayectoria del disco sin mirarlo fijamente o sólo contemplarlo a intervalos discontinuos y breves.

La base geométrica del método de Rawlins se muestra (para un observador situado en el ecuador) en la parte superior de la página siguiente. La primera puesta de sol se ve estando uno tendido en el punto M. En cuanto el último arco visible del disco solar se tienda en el horizonte, hay que levantarse rápidamente y esperar a que ocurra lo mismo, pero ahora desde la altura *h*, que es la altura de sus ojos. El último rayo de la primera observación es tangente a la superficie terrestre en el punto donde el observador se halla y, por tanto, forma un ángulo de 90 grados con la vertical, pero el último rayo de la segunda observación ya no es tangente. Con los ojos del observador situados por encima del nivel del mar, el horizonte del océano forma un ángulo (llamado θ en la figura) con la vertical, su valor es de 90 grados menos la depresión del horizonte (o ángulo de inclinación del horizonte).

El tiempo transcurrido entre la desaparición del último rayo de la primera observación y la desaparición del último de la segunda es proporcional a la depresión del horizonte θ, que es el ángulo que el sol recorre entre los dos ocasos. (En realidad éste es el ángulo girado por la tierra en dicho tiempo.) Si medimos el tiempo en segundos, la proporcionalidad está en el hecho de que a la rotación completa de la tierra le corresponden 86.400 segundos, es decir, 24 horas.

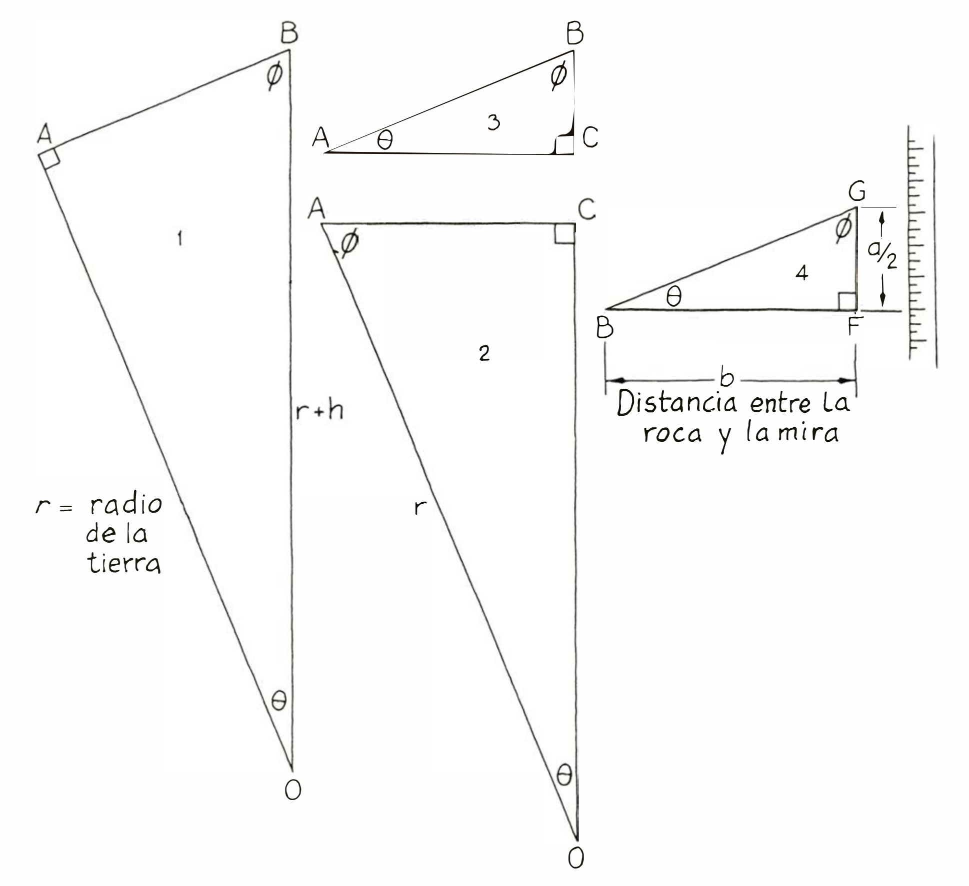
Rawlins admite que el triángulo *QNO* es un triángulo rectángulo y que, por tanto, entre sus lados se puede establecer la relación dada por el teorema de Pitágoras. Uno de sus catetos, *QN*, es aproximadamente igual al radio de la tierra multiplicado por el ángulo que ella recorre entre los dos ocasos. (El cateto es, por tanto, aproximadamente igual al arco entre los puntos *Q* y *M*.) Tomando este dato junto con la relación entre el ángulo y el tiempo transcurrido, Rawlins establece una ecuación que relaciona el tiempo entre los dos ocasos, el radio de la tierra y la altura *h* del nivel de los ojos del observador. (Y supone algo más: la altura de los ojos del observador *h* es despreciable si se la compara con el radio terrestre.) El radio de la tierra en kilómetros es igual a una constante de proporcionalidad, cuyo valor es $3{,}78 \times 10^5$, multiplicada por la altura del nivel de los ojos, *h*, expresada en metros y dividida por el cuadrado del tiempo transcurrido entre los dos ocasos (expresados en segundos).

Si la primera observación no se hace al nivel del mar, la altura a emplear en los cálculos no se deduce tan fácilmente. Supongamos que hacemos la primera observación a la altura normal de los ojos *h*, estando de pie en la playa y que la segunda observación la hacemos después de haber subido un tramo de escalera. La diferencia de altura efectiva para el cálculo es el cuadrado de la diferencia de las raíces cuadradas de las dos alturas de las observaciones. Más adelante daremos un ejemplo de estos cálculos.

Aunque la situación ideal para hacer las observaciones sea un océano o un lago en calma, lo más probable es que haya que contar con el movimiento de las olas. Estas complican las observaciones de dos maneras. Primero, porque las olas de la orilla ocultan la línea del horizonte cuando se está tendido en la playa para hacer la primera observación. Por ello, la primera observación se debe hacer desde un punto más elevado que la altura de las olas. Segundo, porque las olas en la línea del horizonte, que se hallan a unas millas de distancia, elevan éste. Para corregir ese desplazamiento, se aconseja estimar una altura media de las olas y restar dicha altura de cada una de las alturas de las que vamos a efectuar las observaciones. Rawlins, basándose en mediciones oceanográficas, consideró que la altura media de las olas más distantes era de 0,6 metros.

Debe realizarse una corrección sutil con la curvatura de los rayos producida por la refracción en la atmósfera de la tierra. Cuando un rayo de luz pasa de un medio a otro que tenga diferente índice de refracción, su trayectoria se modifica en la superficie de separación de ambos medios. Sólo hay una excepción a lo anterior, que se produce cuando el rayo de luz es perpendicular a la superficie de separación; en ese caso, no se modifica la trayectoria del rayo. La amplitud de la desviación en todas las demás ocasiones depende del ángulo de incidencia del rayo con la superficie de separación y de la diferencia entre los índices de refracción de ambos medios. La dirección de la desviación depende de si el índice de refracción aumenta o disminuye al cruzar la superficie de separación. Si el índice aumenta, el rayo se acerca a la perpendicular y si decrece, se aleja.

Conforme un rayo solar atraviesa la atmósfera terrestre, lo más probable es que se refracte de forma gradual al ir



Triángulos semejantes usados en los cálculos de Gerver

atravesando capas de aire que presenten ligeras diferencias de densidad entre sí y, por tanto, diferentes índices de refracción. La extensión de la desviación es insignificante si la luz tiene una trayectoria casi vertical, pero si el rayo atraviesa la atmósfera con un ángulo grande respecto a la vertical, como ocurre en la puesta del sol, la refracción puede tener un valor significativo.

Si el lector es aficionado a la astronomía probablemente conocerá el efecto de la refracción en las estrellas. Cuando una estrella se encuentra cerca de nuestra vertical, se observa en su posición correcta pues la luz procedente de ella no se refracta demasiado. Sin embargo, cuando una estrella se encuentra justo sobre el horizonte se ve ligeramente fuera de su posición real (el exceso de altura es de alrededor de medio grado en el horizonte), debido a que, cuando el rayo de luz llega a nuestros ojos ha atravesado la atmósfera con un ángulo muy grande respecto a la vertical, lo que hace que venga notablemente refractado. La refracción hace que la estrella parezca tardar más en ponerse debido a que cuando ya está realmente una fracción de grado por debajo del horizonte, la curvatura de sus rayos luminosos hace que su imagen aparezca por encima del horizonte.

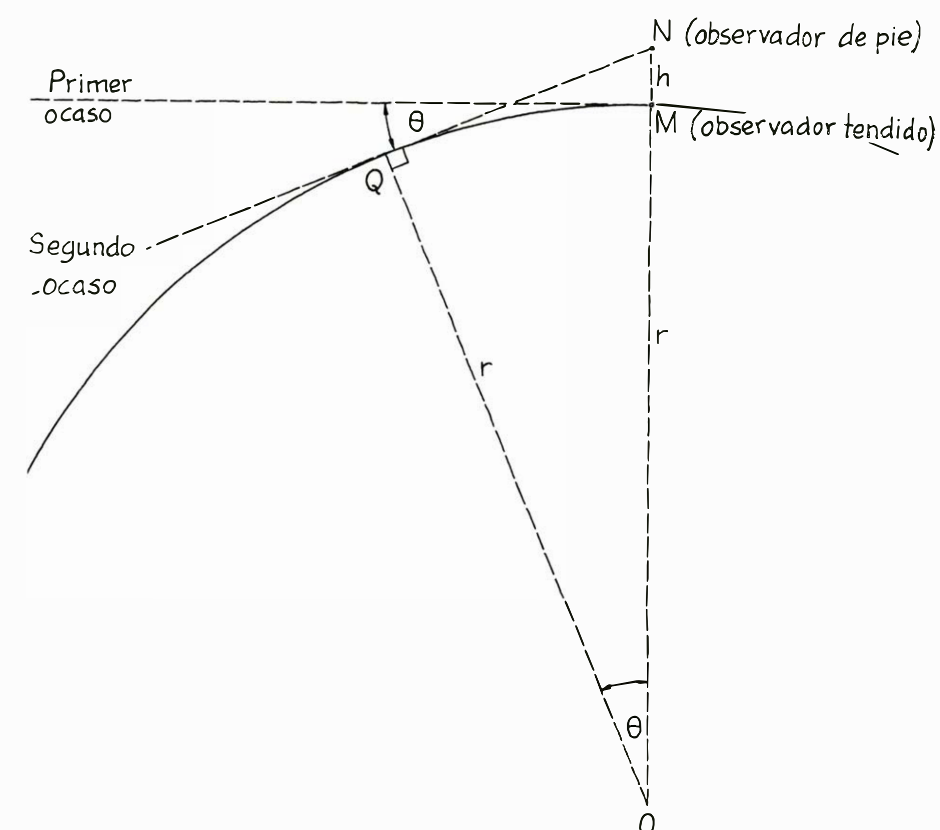
Lo mismo ocurre con el sol, el cual vemos incluso un poco después de ponerse tras el horizonte. Por suceder otro tanto cuando sale, la luz del día dura un poco más de lo que lo haría sin la refracción. Así, debido a que el sol se pone perpendicular al horizonte en el ecuador, podemos contemplarlo por lo menos durante 2,3 minutos más gracias al efecto de refracción. Cerca de los polos el tiempo adicional es mucho más largo porque el sol se pone formando un gran ángulo con la vertical y se pierde por el horizonte más lentamente.

La refracción es fuente de problemas tanto en el método de Gerver como en el de Rawlins. Un rayo de sol que parte del horizonte no viaja en línea recta al observador sino más bien en un círculo que tiene un radio seis veces el de la tierra (un análisis interesante de este efecto aparece en un trabajo clásico de Simon Newcomb, *A Compendium of Spherical Astronomy,* que fue publicado por primera vez en 1906, y reimpreso ahora por Dover).

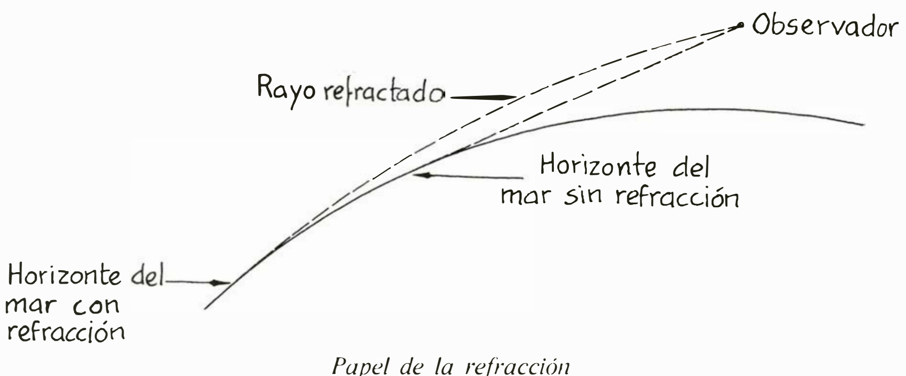
Los efectos de la refracción en los cálculos de Gerver son opuestos a los de Rawlins. En el método de Gerver, la depresión aparente del horizonte se contrae, y, por tanto, disminuye la longitud *a*. Puesto que se necesita esta longitud para varios de los cálculos de proporcionalidad hechos para hallar el radio de la tierra, el error persiste y se llega a deducir un radio que es alrededor del 20 por ciento mayor en virtud de la refracción. En la técnica de Rawlins, la distancia al horizonte del océano aumenta incrementando el ángulo que la tierra debe girar para que se produzca la segunda puesta de sol. Así pues, la refracción aumenta el tiempo entre las puestas de sol y hace disminuir el radio terrestre calculado en alrededor del mismo 20 por ciento.

No se puede contar con el efecto de refracción de una forma segura porque depende de la temperatura y presión de las capas de aire a través de las cuales viajan los rayos de luz. Aquellas magnitudes cambian frecuentemente, no sólo de un punto a otro de la atmósfera, sino también en un mismo sitio. Como una corrección aproximada al efecto de la refracción, Rawlins multiplica su ecuación básica del radio por 1,2. Esta corrección se basa en la idea de que los efectos de la refracción rebajan el radio calculado en aproximadamente un factor 1,2. La corrección reduce de un modo acertado el tiempo medido entre ambas puestas de sol (más exactamente el cuadrado del tiempo) al valor que se obtendría si no hubiera refracción atmosférica.

Sin los refinados procedimientos de Gerver y Rawlins, la única manera de calcular la variación diaria de la refracción es repetir los experimentos muchos días y sacar la media de los radios calcu-



Geometría del método de observar los ocasos de Dennis Rawlins



Papel de la refracción

lados. El factor corrector de 1,2 se verá muy poco modificado. Si la atmósfera está muy estratificada y sus capas tienen un índice de refracción muy diferente, el sol poniente parecerá estar también dividido en capas. El factor corrector de 1,2 será probablemente poco exacto y la medida del radio de la tierra se alejará de la cifra obtenida.

Rawlins hizo otras dos modificaciones a su ecuación básica, relacionadas ambas con el hecho de que en muchos sitios el sol no se pone perpendicularmente al horizonte. Estando en el ecuador, la puesta vertical ideal se puede ver sólo en los equinoccios de primavera y de otoño. En cualquier otra latitud y fecha, las puestas de sol son más lentas porque el sol se oculta formando un ángulo no nulo con la vertical. Rawlins corrige este efecto mediante dos factores que llama *A* y *B*; el primero tiene que ver con la fecha y el segundo con la latitud. *A* es el cuadrado del coseno de la declinación del sol el día del experimento. *B* es el cuadrado del seno de la latitud (con independencia de que sea norte o sur).

El recorrido aparente del sol por el firmamento a lo largo del año determina un plano llamado plano de la eclíptica, que está inclinado 23,44 grados con respecto al plano del ecuador. El sol cruza el plano ecuatorial dos veces al año, en el equinoccio de primavera que tiene lugar sobre el 20 de marzo, y en el otoño, hacia el 23 de septiembre (el momento preciso del equinoccio retrocede cada cuatro años excepto cuando cae en año bisiesto). La declinación del sol es la distancia angular hacia el norte o sur del plano ecuatorial de la tierra. Por ejemplo, en los equinoccios la declinación es cero, en el solsticio de invierno (21 de diciembre) es de unos 23,4 grados hacia el sur y en el solsticio de verano (21 de junio) es de unos 23,4 grados hacia el norte. Se pueden encontrar tablas precisas de la declinación del sol en los anuarios publicados por las principales sociedades astronómicas e institutos relacionados. El lector interesado puede hallar su latitud aproximada en un atlas; con esta latitud y la fecha de las observaciones puede utilizar ya los factores asociados *A* y *B* de la ilustración de esta página. Para usar estos factores correctivos hay que sustraer *B* de *A* y dividir el resultado por la ecuación básica de Rawlins.

Como ejemplo de la fórmula completa, Rawlins proporciona los datos de una de sus observaciones. El 5 de abril de 1978, cronometró una doble puesta de sol en Cove Park en La Jolla, California. La latitud del parque es 32 grados y 51 minutos norte. Rawlins observó la primera puesta de sol a las 6 horas 11 minutos y 55 segundos de la tarde, siendo la altura, desde nivel del mar a sus ojos, de unos 1,72 metros. Después de subir a una escalera hizo su segunda observación a las 6 horas 12 minutos y 15 segundos de la tarde, a una altura de 8,95 metros sobre el nivel del mar. El tiempo registrado por su cronómetro desde los últimos rayos de la primera puesta y los de la segunda fue de 19,6 segundos.

Rawlins calculó luego la altura añadida al horizonte del océano por las olas distantes en 0,6 metros. Sustrajo esta cifra de las dos alturas de observación, que pasaron a ser 1,12 y 8,35 metros. Para obtener la diferencia de altura efectiva necesaria para el cálculo, extrajo la raíz cuadrada de ambas cifras, halló la diferencia entre ellas y elevó al cuadrado el resultado, obteniendo un valor final de 3,55 metros.

Elevando al cuadrado el tiempo, dividiendo el resultado por la constante de proporcionalidad de 3,78 x 10^5 y multiplicando por la diferencia de altura efectiva de 3,35 metros, obtuvo el resultado de 3300 kilómetros como medida del radio terrestre, cifra que resultaba ser un 48 por ciento demasiado baja, pero los factores relacionados con la fecha, la latitud y la refracción no se habían tenido aún en cuenta.

Interpolando a partir de las tablas, Rawlins descubrió que el día 5 de abril tenía un factor *A* de 0,988. El factor *B* para su latitud era aproximadamente 0,295. *A* menos *B* era 0,693. Dividió esta cifra por el resultado anterior y obtuvo un radio de 4800 kilómetros, un resultado aún bajo en un 25 por ciento. Para utilizar el dato de la refracción multiplicó este resultado por 1,2 y así obtuvo el valor final de su radio: 5700 kilómetros. Esta cifra es inferior a la oficial en un 10 por ciento.

Si el lector vive cerca de un océano o de un lago grande, quizá quiera repetir el experimento de Gerver o el de Rawlins, recogiendo suficientes datos que le permitan encontrar un valor medio para el radio de la tierra. Dado que la refracción atmosférica afecta a ambos experimentos en direcciones opuestas, si se realizan ambos y se saca la media de los resultados, se pueden casi eliminar los efectos de la refracción. A mí me resulta impresionante que se pueda llegar a una cifra muy cercana al radio real de la tierra con medidas aproximadas y sin más equipo que una regla o un reloj y aún es más impresionante que se pueda llegar a resultados que difieran sólo entre un 10 y un 20 por ciento de las medidas oficiales.

Hace mucho tiempo que el hombre se dio cuenta de que la tierra era redonda debido a ciertas pistas: las estrellas tenían posiciones muy diferentes vistas desde lugares separados hacia el norte o hacia el sur, los barcos desaparecían en

Día del año (mes/día)				A	Latitud	B
3/20	9/22	9/22	3/20	1	0°	0
3/27	9/15	9/30	3/13	.9975	5°	.0076
4/4	9/7	10/7	3/6	.99	10°	.030
4/10	9/1	10/14	2/27	.98	15°	.067
4/15	8/27	10/19	2/22	.97	20°	.117
4/20	8/22	10/23	2/18	.96	25°	.179
4/24	8/18	10/27	2/14	.95	30°	.250
4/28	8/14	10/31	2/10	.94	35°	.329
5/1	8/10	11/3	2/6	.93	40°	.413
5/5	8/7	11/7	2/3	.92	45°	.500
5/9	8/3	11/11	1/30	.91	50°	.587
5/13	7/30	11/14	1/27	.90	55°	.671
5/17	7/26	11/18	1/23	.89	60°	.750
5/21	7/22	11/22	1/19	.88	65°	.821
5/26	7/17	11/27	1/14	.87	70°	.883
5/31	7/12	12/2	1/9	.86	75°	.933
6/7	7/5	12/8	1/3	.85	80°	.970
6/12	6/29	12/13	12/29	.845	85°	.9924
6/21	6/21	12/21	12/21	.8417		

Los factores **A** *y* **B** *usados por Rawlins*

el horizonte, la sombra de la tierra en la luna durante un eclipse lunar era redonda. Sin embargo, para muchos hombres fue difícil aceptar la idea de una tierra redonda.

Narra la leyenda que un sabio griego, Eratóstenes, que trabajaba en Alejandría, realizó las primeras mediciones del tamaño de la tierra hace unos 2200 años. Para ello utilizó un método muy simple (casi rudo), pero sorprendentemente obtuvo buenos resultados. En Siene (ahora Asuan) advirtió (o le hicieron caer en la cuenta) que por la tarde, en el solsticio de verano, los rayos solares entraban en un pozo verticalmente. Esto le llevó a pensar que él podría calcular la circunferencia de la tierra si supiera el ángulo (desde la vertical) de los rayos de sol al mediodía en la misma fecha en Alejandría, que estaba al norte de Siene y situada casi en el mismo meridiano.

Siempre según la leyenda, las mediciones de Eratóstenes en Alejandría demostraron que los rayos se desviaban de la vertical en una cincuentava parte del círculo (7,2 grados). Dado que los rayos de Alejandría eran paralelos a los de Siene (debió observar que la distancia al sol sobrepasaba mucho al radio de la tierra), la diferencia de ángulo tenía que ser el resultado de la curvatura de la tierra. El arco entre las dos ciudades a lo largo de la superficie terrestre ocupa un ángulo de 7,2 grados con respecto al centro de la tierra.

Así pues, para hallar la circunferencia de la tierra lo único que tenía que hacer Eratóstenes era deducir la relación de este ángulo y 360 grados e igualarla con la razón entre la longitud del arco que separa ambas ciudades y la circunferencia de la tierra. Sin embargo, no fue fácil encontrar la distancia entre una y otra ciudad. Dando pruebas de un gran ingenio, Eratóstenes midió la distancia tomando como base el tiempo que tardaban los camellos en ir de una ciudad a otra, que eran 50 días aproximadamente. Sabiendo qué recorrido hacía un camello en un día, pudo calibrar la distancia entre Alejandría y Siene y así calcular la circunferencia de la tierra. Su resultado (en medidas modernas) fue de 46.250 kilómetros, cifra que excede en un 16 por ciento el valor real.

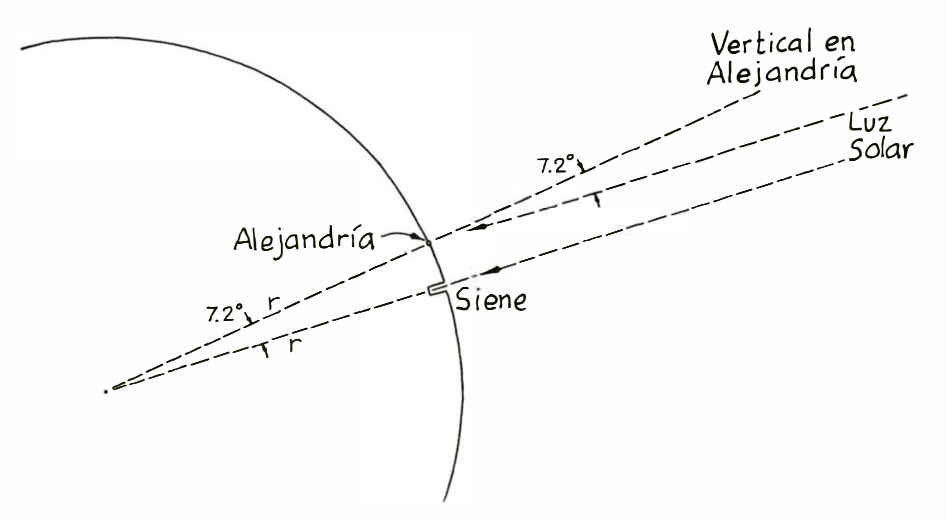
Siguiendo la historia, la siguiente medición de la circunferencia de la tierra ocurrió en el año 85 antes de Cristo, por Poseidón de Apamea, quien usó la distancia entre Rodas y Alejandría de la misma forma que Eratóstenes lo hiciera con Alejandría y Siene. El medio empleado por Poseidón para medir la distancia entre ambas ciudades fue un barco de línea regular que las unía. Poseidón midió el ángulo observando la estrella Canopus; cuando estaba en el horizonte de Rodas, se hallaba 7,5 grados por encima del horizonte de Alejandría. Se dice que Poseidón llegó a calcular la circunferencia de la tierra haciendo el mismo tipo de cálculos que Eratóstenes. Los cálculos modernos afirman que sus mediciones eran erróneas en una sexta parte o alrededor del 17 por ciento.

Posteriormente, Abdullah al-Hamun realizó otra medición, y sus resultados se alejaban de la cifra real en sólo un 3,6 por ciento. Dispuso de datos astronómicos muy exactos para hallar el ángulo que necesitaba e incluso llegó a medir la distancia de su arco colocando señales en los extremos del mismo.

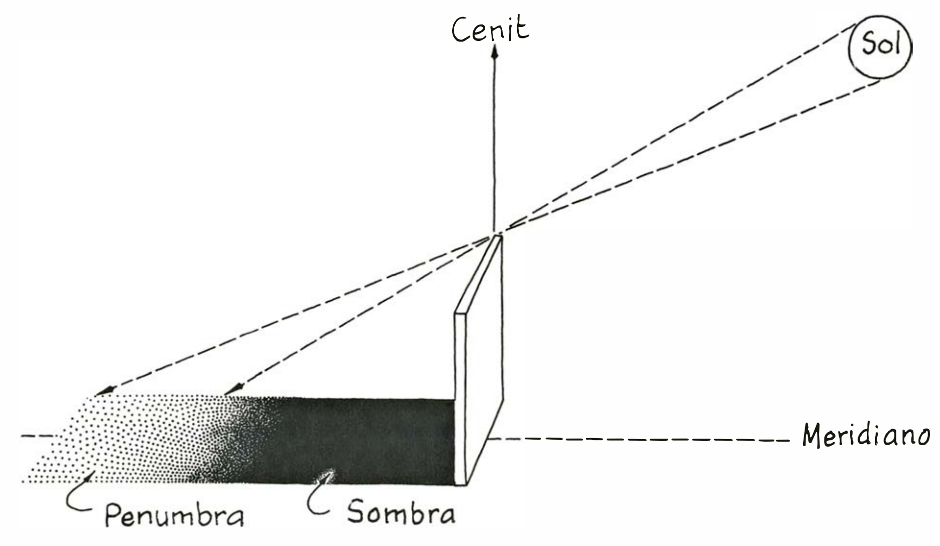
Las técnicas modernas son, por supuesto, más exactas. Se sirven de satélites artificiales, rayos láser y reflectores en la luna, con lo cual el tamaño de la tierra es perfectamente conocido así como el hecho de que no es totalmente esférica. Han transcurrido muchos siglos desde la época de los antiguos griegos que creían que la tierra era plana y que estaba sostenida por cuatro enormes elefantes que a su vez se sostenían sobre una inmensa tortuga. A nosotros no nos preocupa ya el problema filosófico de sobre qué se asienta la tortuga.

A lo mejor el lector quiere medir el radio de la tierra con la técnica atribuida a Eratóstenes. En este caso es aconsejable elegir dos ciudades cuyos respectivos meridianos estén próximos, aunque esta exigencia no es esencial. Las ciudades deben estar referenciadas por una medida que resulte práctica. (Esta podrá encontrarse en un mapa o en una tabla de distancias.)

Para medir la orientación de los rayos solares en ambas ciudades hace falta que dos amigos, uno en cada ciudad, lleven a cabo el experimento el mismo día. Ambos deberán levantar un gnomon, un poste rectangular vertical con una cara perpendicular al meridiano. Cuando el sol alcance su cenit, se mide la longitud de la sombra producida por



Geometría del método atribuido a Eratóstenes



Método del gnomon para determinar la orientación de los rayos solares

el gnomon (será la más corta entonces). La tangente de la distancia angular del sol desde el cenit es igual a la longitud de la sombra dividida por la altura del gnomon. Aunque se podría calcular entonces el ángulo con una tabla de funciones trigonométricas, mi sugerencia es hacerlo con una calculadora de bolsillo para evitar el problema de interpolación entre los valores de la tabla. El ángulo entre la vertical y los rayos solares es, por supuesto, el mismo.

Una vez tomadas las medidas, hay que restarles los ángulos obtenidos por ambos observadores. Después, siguiendo el procedimiento de Eratóstenes se divide la diferencia angular por 360 grados y se multiplica por la longitud del arco entre las dos ciudades. El resultado es la circunferencia de la tierra.

El inconveniente de este método es que el extremo más alejado de la sombra no es muy nítido, ya que va pasando desde la total brillantez a través de una zona de iluminación parcial (la penumbra) hacia un área de máxima obscuridad (la sombra). La penumbra está iluminada parcialmente, porque la plancha bloquea los rayos que se originan desde la parte más baja del sol pero no bloquea los de la parte superior. Dentro de la sombra todos los rayos directos del sol quedan bloqueados por la plancha.

Para hallar la distancia angular entre el cenit y el centro del sol se calcula la sombra hasta el punto central de la penumbra. Esta medida es difícil de realizar con precisión e introduce inevitablemente un error que puede llegar a ser de un cuarto de grado si se usa cualquier borde de la penumbra en la determinación del ángulo de los rayos de sol. La penumbra es 0,5 grados de ancho porque el sol subtiende un ángulo de 0,5 grados en el campo de visión del observador. El error inherente a la localización del centro de la penumbra hace del gnomon un instrumento impreciso para calcular el tamaño de la tierra.

El incremento puede salvarse sustituyendo el gnomon por otro aparato. Verbigracia: montando un perno horizontal en el centro de un círculo graduado colocado de forma que su plano sea vertical y esté alineado perpendicularmente al meridiano. La sombra producida por el perno tendrá zonas de penumbra a lo largo de la longitud de la sombra. Por ser simétricas en los dos lados de la sombra, es fácil encontrar el centro de la sombra y leer su posición en el círculo graduado. Cuando el círculo está bien orientado, la lectura resultante será la distancia angular del sol desde el cenit y, por tanto, el ángulo de los rayos incidentes sin ningún error sistemático introducido por las zonas de penumbra.

Libros

Un nuevo enfoque de la química orgánica, radiobiología médica y Ramón y Cajal

R. Pérez Alvarez-Osorio, J. M. Arranz y M. Durfort

QUIMICA ORGANICA EXPERIMENTAL, por Donald L. Pavia, Gary M. Lampman y George S. Kriz, Jr. Eunibar, Editorial Universitaria de Barcelona, 1978. En el inicio de la formación de un químico orgánico se requieren fundamentalmente dos tipos de libros: el libro de texto, en el que se exponen los aspectos teóricos de la química orgánica, se describen las relaciones estructurales entre los compuestos orgánicos y se detallan sus reacciones, los mecanismos que siguen las mismas, el campo de aplicación de ellas, etcétera. Junto al texto se ha considerado siempre indispensable el libro de prácticas que enseña a operar en el laboratorio, a manejar los métodos generales de trabajo, a extraer los productos de sus mezclas naturales o artificiales y a iniciarse en el arte de la síntesis, así como a utilizar las técnicas analíticas más elementales. Aunque, en épocas relativamente recientes, se ha considerado conveniente incorporar a la bibliografía del aprendiz de química orgánica el libro de problemas, el manual de nomenclatura, las tablas espectroscópicas, el folleto programado y hasta cierto número de monografías, el texto teórico y el manual de prácticas constituyen aún los dos pilares básicos para que el químico orgánico pueda comenzar su andadura. El atractivo que el estudiante pueda encontrar en las obras de este tipo que utiliza en sus primeros años será decisivo para la forja de su vocación.

Estas simples consideraciones justifican la exigencia más rígida no ya en la redacción y traducción de ambos tipos de obras, sino también en los aspectos materiales de presentación de ellas, ilustraciones, impresión, distribución del texto, etc., aspectos, por cierto, muy bien cuidados en la actualidad. Como es lógico, la selección de un libro de texto o de prácticas constituye una responsabilidad seria por parte del profesor, que, al hacerla, ha de tener en cuenta, incluso, las cambiantes características y aficiones del material humano con que trabaja.

Nos parece que los textos teóricos han experimentado, hace tiempo, esta evolución "cara al público", dicho sea sin ningún matiz peyorativo. La presentación pictórica de las teorías estructurales, aunque a veces conduzca a simplificaciones excesivas, es una de las características más relevantes. Representación que se extiende ya a los estados de transición de las reacciones y pisa atrevidamente el terreno dinámico de la reactividad.

Sin embargo, los textos prácticos conservaban, en general, una línea más tradicional y clásica. Su contenido y ordenación no variaban mucho de los utilizados a principio de siglo: operaciones generales, habiéndose lógicamente sumado a las clásicas algunas más recientes, en particular técnicas cromatográficas; a continuación, síntesis y reacciones, ordenadas según el esquema habitual: hidrocarburos, derivados halogenados, alcoholes, etc. o con variantes encaminadas a graduar las dificultades de manipulación; alguna práctica analítica de tipo espectroscópico completa a veces el tratado.

¿En qué idea directriz podría basarse la modificación de este esquema? Una muy feliz es la que desarrollan los profesores D. L. Pavia, G. M. Lampman y G. S. Kriz en su "Química orgánica experimental" cuya versión española presenta la Editorial Universitaria de Barcelona. Digamos, en primer lugar, que al libro, a juzgar por la lectura de su primer volumen y por el contenido que se anuncia del segundo, no le falta nada que pueda encontrarse en otras obras de finalidad análoga. Pero contiene mucho más que la mayoría de éstas y posee un enfoque original y ciertamente poco frecuente.

La idea que ha guiado a los Profesores Pavia, Lampman y Kriz es la de elegir como productos orgánicos a preparar o aislar aquellos que tienen interés práctico por sus relaciones con la biología o con procesos de interés industrial. En lugar de omitir las preparaciones de interés puramente académico, se ensamblan éstas con otras que llevan a productos de uso frecuente en la vida diaria. Se estimula así a los estudiantes que no tienen a la química como objetivo principal de sus estudios y que siguen los de Farmacia, Biología o Medicina.

No faltan, como decimos, los métodos de síntesis y el estudio de la reactividad de productos clásicos como los aromáticos nitrobenceno, anilina y ácido benzoico: asimismo, los halogenuros de alquilo y alquenos. Pero estas preparaciones se relacionan con temas de interés general, como los estudios cinéticos, y buscan la relación con problemas biológicos o industriales. Un ejemplo típico de la particular manera con que abordan los autores la preparación práctica de los estudiantes, lo constituyen las prácticas 38, 39 y 40. Se trata de las clásicas preparaciones de benzoína, bencilo y ácido bencílico, pero la primera se hace en presencia de tiamina como catalizador, lo que les permite estudiar el carácter de coenzima de este compuesto.

La originalidad de la obra salta pronto a la vista, cuando el lector comprueba que la primera práctica es la presentación de la aspirina a la que siguen las de otros analgésicos y el análisis de una preparación desconocida de este tipo de fármacos. La extracción de productos naturales adquiere en esta obra mucha mayor extensión que en otros tratados prácticos de química orgánica: la extracción de la nicotina del tabaco, de la cafeína del café y del té y del colesterol de los cálculos biliares figuran entre las primeras prácticas. El carácter "aplicado" de la obra queda de manifiesto si reseñamos que en ella se describen métodos de análisis de grasas y de preparación de jabones y detergentes, procedimientos de obtención de varios polímeros, el aislamiento de cierto número de terpenos y el estudio de varios aceites esenciales, numerosas preparaciones de interés farmacológico, como sulfamidas, anestésicos y barbitúricos. Los aspectos más recientes están representados, por ejemplo, por la síntesis de una feromona.

Las técnicas generales se detallan en el volumen II, aún no publicado en nuestro país. A ellas se hace continua referencia y es indudable que el estudiante ha de manejar ambos volúmenes simultáneamente para obtener el máximo fruto de esta obra. Hay, además, que destacar muy particularmente, una serie de introducciones generales, en las que figuran los consejos sobre la manera de actuar en el laboratorio, que aquí se amplían con detalles que no suelen encontrarse en otras obras, por ejemplo, los peligros del uso inadecuado de los disolventes, con citas sobre las enfermedades que pueden producir éstos, el calzado que debe llevarse durante el trabajo práctico, etc.

Los autores justifican la extensión de su obra declarando que se trata de tres libros en uno: las prácticas, las técnicas

y las "instrucciones". Estas últimas constituyen, entre todos los aspectos originales del libro, el que encierra mayor novedad. Son breves artículos, de dos a cuatro páginas, conteniendo una exposición de un tema (la aspirina, la nicotina, la química de la visión, las feromonas, la química de la leche, etc.) magistralmente resumido y documentado, en cada caso, con una media docena de citas. Se dan sabrosas noticias históricas y se plantea cada tema con gran énfasis en su trascendencia en la vida humana. La lectura de esta introducción constituye un auténtico solaz y una forma de comunicar información realmente acertada. El estudiante capta un problema y en la práctica o prácticas que siguen comprueba cómo él mismo puede resolverlo. Se da, así, un hábil ensamblaje del porqué y el cómo de la química orgánica que produce un impacto profundo en el estudiante. El profesor obtendrá, sin duda, de estas introducciones una serie de datos curiosos que contribuirán a hacer amenas sus disertaciones.

La presentación de la obra es muy atractiva, la impresión muy clara, los espectros y figuras nítidamente impresos.

Podríamos objetar algunos puntos concretos de la traducción. Parece que nuestros traductores no se han puestode acuerdo sobre el uso del infinitivo o de la forma reflexiva en la descripción de las prácticas. Los traductores de esta obra prefieren la primera solución, lo que da cierta dureza a la lectura de la parte correspondiente. Por nuestra parte, "se ponen 25 g de..." es preferible a "poner 25 g de...". Hay frases cuya corrección castellana resulta al menos dudosa: "Es por ello que Duisberg ideó..." (pág. 26); creemos que bastaría "Por ello Duisberg ideó...". En la página 59 "Como que posee ocho centros asimétricos y son posibles, por tanto, 2^8 o 256 estereoisómeros, se precisaron otros 23 años hasta que se pudo elucidar su estructura tridimensional" es indudable que sigue con demasiada rigidez el texto original. Otras muestras podrían citarse. Pero el conjunto resulta claro y de grata lectura. En resumen, una valiosa adición a la bibliografía químico-orgánica que será preciada y utilizada por profesores y estudiantes. (R. P. A-O.)

RADIOBIOLOGIA MEDICA, por Elizabeth Latorre Travis. Editorial AC, Madrid, 263 págs. (1979). Es la primera edición en castellano de la obra "Primer of Medical Radiobiology" publicada por Year Book Medical Publishers, Inc., Chicago, III, U.S.A. Es intelectualmente imprescindible y fundamental para conseguir una correcta preparación técnica que el radiólogo conozca y procure relacionar los más diferentes aspectos de la radiología: la radiobiología, las técnicas radiológicas y la radioprotección. Esto mismo, quizás a menor escala, puede extenderse a todo profesional que maneje radiaciones ionizantes y a los estudiantes que dediquen sus esfuerzos a conocer las distintas posibilidades que se ofrecen en este campo.

Esta obra se propone presentar la forma como actúan las radiaciones sobre el organismo y destacar la importancia que tienen esos mecanismos de actuación como productores de efectos en las diferentes regiones donde intervengan. El contenido del texto se divide en diez capítulos. Los primeros sirven de introducción, y exponen los conocimientos básicos de biología celular necesarios para comprender las interacciones biológicas de la radiación. Al exponer la radiosensibilidad en general se indican los hechos fundamentales para permitir una explicación de la acción de las radiaciones sobre los sistemas biológicos.

Los capítulos centrales describen la respuesta a las radiaciones a nivel celular, a nivel sistémico y por último a nivel del organismo en conjunto. Durante mucho tiempo se ha sabido que las radiaciones ionizantes dañan tejidos del cuerpo, si bien la relación entre lesión y dosis siempre ha constituido un punto difícil de establecer en los seres humanos por el peligro que supone el extrapolar a nuestra especie los resultados experimentales obtenidos en animales de laboratorio.

Se han realizado investigaciones fundamentales sobre las unidades íntimas de los tejidos, las células, todas las cuales sufren daño directo por la radiación. En el texto se recogen los diversos factores físicos, químicos y biológicos que influyen en la respuesta celular a la misma. El efecto producido por la radiación en el cuerpo depende del área irradiada. La exposición del cuerpo entero implica daño a todos los tejidos radiosensibles y, además, si la dosis es grande, los efectos producidos serán permanentes; mientras que si es pequeña, el tejido podrá recuperarse. Sobre la piel, dosis grandes de radiación producen enrojecimiento y pueden dar lugar eventualmente a su destrucción o a un cáncer radioinducido. Pueden destruir los órganos de reproducción, haciendo al sujeto estéril, o dañar el órgano de la visión produciendo opacidad del cristalino. Estos efectos y otros son expuestos detalladamente en los capítulos dedicados a la respuesta de los diferentes tejidos y sistemas orgánicos a las radiaciones.

El bloque fundamental de la obra está compuesto por los capítulos de radiobiología clínica que abarcan sucesivamente las técnicas de radiodiagnóstico, medicina nuclear y radioterapia. El radiodiagnóstico, cada vez más empleado en la práctica médica, somete tanto al enfermo como al personal especializado a una cantidad de radiación que será necesario medir y valorar. En el texto se analizan los riesgos del empleo de esta técnica diagnóstica, insistiendo en los efectos de la irradiación fetal y en la importancia de reducir la exposición del paciente siempre que sea posible.

La medicina nuclear, en los últimos años, ha experimentado grandes progresos en el campo diagnóstico. La introducción de la gammacámara que permite una rápida producción de imágenes y la utilización de radioisótopos de vida media corta son un ejemplo de los factores que han contribuido a ese desarrollo, así como también el descubrimiento de nuevos compuestos y de coloides para fines específicos de escintigrafía que han revolucionado por completo el campo de aplicación de la gammagrafía. En el texto, el estudio de la exposición a las radiaciones en las técnicas de medicina nuclear se amplía además con una somera descripción de la utilización terapéutica de los radionúclidos, aunque una de las tendencias bastante aceptadas actualmente es potenciar los protocolos de tratamiento quimioterápico, relegando en parte los tratamientos con agentes radiactivos.

La base de la radioterapia es la destrucción selectiva de tejidos. Las células neoplásicas, debido a la modificación de su patrón mitótico, son casi siempre más sensibles que sus células progenitoras. La diferencia de radiovulnerabilidad entre tejidos sanos y cancerosos determinará en gran parte el que se pueda erradicar una determinada lesión. Es necesario comprender el significado de los términos radiosensibilidad y radiocurabilidad, que, aunque se asocian en ocasiones, no son sinónimos. Así, mientras las células malignas de un linfosarcoma generalizado son muy radiosensibles, ese linfosarcoma es muy raro que sea radiocurable; por el contrario, las células malignas de un carcinoma de células basales de la piel, que son bastante radiorresistentes, no impiden el que este tipo de tumor sea con frecuencia radiocurable.

Estas cuestiones se afrontan en la obra y en ella se trata de describir los principios básicos que rigen la radioterapia de las enfermedades malignas cancerosas con el objeto de obtener cada vez mejores resultados en los tratamientos. Asimismo, se habla de la importancia del fraccionamiento de las dosis, de elevado interés para la protección del tejido conjuntivo, y del concepto de tolerancia, persiguiendo siempre conservar la inte-

gridad del tejido normal. El glosario final presenta de forma didáctica y resumida el significado de los términos fundamentales en la especialidad.

Se trata, en conjunto, de un texto coherente por su contenido temático y recomendable para todo profesional que maneje las radiaciones ionizantes, ya que encontrará en él información clara de los riesgos biológicos que ocasiona su arma de trabajo y le facilitará el comprender que el temor con que se encuentra envuelta su profesión no tiene razón de ser si las radiaciones se utilizan de una forma racional, conociéndolas a fondo y ajustando su empleo sólo a los casos en que sea seguro obtener algún efecto positivo. (J. M. A.)

SANTIAGO RAMON Y CAJAL, por Agustín Albarracín. Editorial Labor, Barcelona 1978. Cualquier momento es oportuno para aparecer una obra dedicada a un hombre ilustre, sea cual fuere su especialidad. Salvador de Madariaga, en una de las últimas entrevistas concedidas a la prensa, lamentaba el poco énfasis con que los españoles trataban a sus más preclaras figuras, señalando que la internacionalidad alcanzada por Napoleón podría haber sido igualmente lograda y con toda justicia por personalidades de la índole de Hernán Cortés.

En este sentido cabe destacar que Santiago Ramón y Cajal ya en vida logró que se reconociera su talento, que se valorara con justicia su obra, primero en el extranjero y posteriormente en su país, habida cuenta de que los grandes maestros de la histología estaban en Alemania y Francia. Por lo que fue necesario que Cajal publicase, con grandes dificultades económicas, sus trabajos y asistiese a los grandes congresos poniéndose en contacto con las personalidades más relevantes de la histología, venciendo la, hasta cierto punto, lógica prevención sobre la calidad e interés que podía aportar a la ciencia los trabajos llevados a cabo por un joven médico español. Aunque España era conocida como país de ilustres literatos y magníficos pintores, no había dado hasta el momento figuras relevantes y de proyección hacia el exterior en el campo de las ciencias.

Al respecto, en la introducción de la obra que comentamos, escrita por Laín Entralgo, se recoge el comentario de Ortega y Gasset en 1908, ante la Asamblea para el Progreso de las Ciencias: "Somos culturalmente insolventes, arrastramos una deuda secular del espíritu... El caso Cajal y mucho más el caso Hinojosa no pueden significar un orgullo para nuestro país, son más bien una vergüenza, porque son una casualidad".

La presente obra es el exponente de una labor cuidada, matizada en todos sus detalles. Los autores, Laín Entralgo y Albarracín Teulón, ya habían publicado en 1967 un libro "Nuestro Cajal" en la colección –Nombres que son Historia– siguiendo los cánones habituales en las biografías, una ordenación cronológica perfecta, etc. La presente obra difiere de la anterior y de otras muchas por estar elaborada principalmente a base de descripciones, de situaciones y de hechos, relatados por el propio Cajal, haciendo a la vez un exhaustivo estudio del contexto histórico en que se van desarrollando los acontecimientos.

No hallamos las muy conocidas aventuras y travesuras de Cajal niño, en cambio con breves princeladas aparecen aspectos curiosos del Cajal adolescente y joven, como son su preocupación por la complexión física, su sensibilidad hacia los encantos femeninos, que perduraría en el transcurso del tiempo, los tiernos comentarios hacia su mujer...

De la mano de Albarracín Teulón nos situamos en Valencia, cuando, tras los estudios de medicina cursados en la Facultad de Zaragoza, gana la cátedra de anatomía de la Facultad de Valencia. Tras una corta estancia en Madrid, por razones burocráticas, Simarro le da a conocer la técnica seguida por el italiano Camilo Golgi, para la tinción específica de las células nerviosas, a base de la impregnación cromo-argéntica, siendo éste un momento trascendental para los posteriores trabajos de Cajal. De regreso a Valencia, empieza a ensayar dicha técnica sobre materiales de distinta textura y a la vez introduce modificaciones al método original.

Su toma de posesión en la Facultad de Medicina de Barcelona abre una nueva etapa en su actividad y, según él mismo escribe, el año 1888 es un año clave en el progreso de su investigación. Es particularmente recomendable la lectura de sus impresiones al observar una preparación de tejido nervioso impregnado con la modificación del método cromo-argéntico de Golgi (pág. 102), considerando que el entusiasmo que manifiesta puede ser solamente apreciado en toda su magnitud por los que trabajan en algún campo experimental y particularmente por los histólogos.

Es importante el año 1889, en que Cajal asiste y participa activamente en la reunión Anual de la Sociedad de Anatomía Alemana, entra en contacto con las más preclaras figuras de la histología del momento y entabla conocimiento personal con Kolliker, quien evidencia una gran humanidad cuando, tras escuchar a Cajal y observar las preparaciones microscópicas que había llevado consigo a fin de demostrar que las imágenes que

había presentado eran reales y no artefactos de técnica, le escribió: "Los resultados obtenidos por usted son tan bellos que pienso emprender inmediatamente, ajustándome a la técnica de usted, una serie de trabajos de confirmación. Le he descubierto a usted, y deseo divulgar en Alemania mi descubrimiento" (página 104).

Su asentamiento definitivo en la cátedra de histología normal y de anatomía patológica en la Facultad de Medicina de Madrid (1892), a la muerte del ilustre Maestre de San Juan, es la iniciación de una etapa de estabilidad, de relativa tranquilidad, de trabajo continuado y fecundo, de la creación de un equipo de discípulos primero y colaboradores después. La creación del Laboratorio de Investigaciones Biológicas, así como la edición de la publicación "Trabajos del Laboratorio" (1902), le llenan de noble satisfacción.

Es particularmente interesante, aunque muy breve, uno de los últimos capítulos de la obra ("En sus manos está el porvenir de la ciencia española") en el que tras unas líneas de presentación se dan a conocer los histólogos de la época que estuvieron más relacionados con Cajal, y entablan un diálogo con el maestro, dejándose entrever algún que otro empecinamiento de Cajal que lo hacen todavía más humano. Desfilan Tello, Achúcarro, Río Hortega y Castro, su sucesor en la cátedra y discípulo más querido.

Deprime comprobar la premonición que tuvo Cajal al escribir, a sus 82 años, poco antes de morir, a Ortega y Gasset una carta en la que el histólogo señala su preocupación ante la imposibilidad de no poder reeditar su obra "Textura del sistema nervioso del hombre y de los vertebrados", cuya primera parte fue editada en 1899 y la segunda en 1904, agotándose rápidamente; en 1909 se publica la traducción de la obra en francés por el histólogo Azoulay. Agotada desde hacía diez años y en la que, a pesar de su avanzada edad, estaba trabajando, preparando nuevos esquemas, seleccionando fotografías inéditas, ampliando determinadas partes, etcétera. Cajal tuvo el presentimiento que no vería cumplido este último deseo. Con motivo del homenaje que se le tributó en el I Centenario de su nacimiento, se reeditó, en 1952, dicha obra, si bien tras acaloradas discusiones entre los que intervinieron a favor de publicarla intacta y los partidarios de introducir sus últimas anotaciones personales, apareció la reedición de la traducción francesa del año 1909.

A través de los textos, muy bien escogidos por Albarracín, llegamos al año 1934, año de la muerte del ilustre histólogo español, a la enviable edad de 82 años, tras una vida fecunda y haber legado a la ciencia numerosos trabajos de valor indiscutible y que no han perdido vigencia, siendo esto lo mejor que se puede decir de la labor de un hombre. Efectivamente, trabajos llevados a cabo con el microscopio electrónico estos últimos años, sobre materiales estudiados por Cajal con sus magníficas técnicas de impregnación, confirman la validez en todos los casos de las interpretaciones dadas por él.

La iconografía del libro que comentamos es completa y oportuna. Seleccionada con tanto acierto que gráficamente se nos transmiten las aficiones predilectas de Cajal; su gran pasión por la fotografía, no sólo la microscópica, sino la macroscópica, queda evidenciada en la profusión de retratos familiares, autorretratos, algunos realmente anecdóticos, como los dos que nos ofrecen la imagen de un Cajal atlético, para muchos desconocido, cuya musculatura, como dice él mismo, ha sido ligeramente retocada para hacerla más prominente. La reproducción de su correspondencia epistolar, de portadas de trabajos, notas de laboratorio, telegramas, cartas de felicitación con motivo de la concesión del Premio Nobel de Fisiología y Medicina de 1906, las remitidas a raíz de su jubilación, etc., hacen que esta obra palpite de humanidad, que sea algo realmente vivo.

Consideramos que la relación bibliográfica es insuficiente, si bien los autores ya señalan que se trata de un resumen. A los interesados en la obra científica de Cajal, nos permitimos recomendarles la consulta del tomo XXX de los Trabajos del Laboratorio de Investigaciones Biológicas del año 1935, en que la tercera. parte de la obra es una relación de las publicaciones de don Santiago, comentadas por él mismo.

Recomendamos esta obra a toda persona amante de las biografías y de los libros en general, por su contenido y por su magnífica y cuidada presentación, creyendo sinceramente que debería ser conocida por todo el profesorado universitario, ya que en su lectura se palpan una serie de problemas que han sido, son y serán. Así en la página 156, recordando su estancia en Londres y las impresiones de su paso por Cambridge, Cajal escribe: "¡Qué pena, otra vez, recordar el panorama de nuestros estudiantes aislados, de nuestros profesores encastillados en sus cátedras como lechuzas en campanario, desconociéndose entre sí y ajenos por completo a los nobles anhelos de una colaboración orgánica, como si no formaran parte de un mismo cuerpo ni conspiran al mismo fin!" (M. D.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

MONOXIDO DE CARBONO EN LA GALAXIA

THE MILKY WAY. Bart J. Bok y Priscilla F. Bok. Harvard University Press, 1974.

MAN DISCOVERS THE GALAXIES. Richard Berendzen. Richard Hart y Daniel Seeley. Science History Publications, 1976.

THE LARGE-SCALE CHARACTERISTICS OF THE GALAXY. Proceedings of the International Astronomical Union. Symposium n.º 84, dirigido por W. B. Burton. D. Reidel Publishing Company. 1979.

EL SPIN DEL PROTON

ELASTIC PROTON–PROTON SCATTERING AT 90º AND STRUCTURE WITHIN THE PROTON. C. W. Akerlof, R. H. Hieber, A. D. Krisch, K. W. Edwards, L. G. Ratner y K. Ruddick en *Physical Review,* vol. 159, n.º 5, págs. 1138-1149; julio, 1967.

ACCELERATION OF POLARIZED PROTONS TO 8.5 GEV/C. T. Khoe, R. L. Kustom, R. L. Martin, E. F. Parker, C. W. Potts, L. G. Ratner, R. E. Timm, A. D. Krisch, J. B. Roberts y J. R. O'Fallon en *Particle Accelerators,* vol. 6, n.º 4, págs. 213-236; 1975.

SPIN-SPIN INTERACTIONS IN HIGH $P_\perp^2$ ELASTIC P-P SCATTERING. J. R. O'Fallon, L. G. Ratner, P. F. Schultz, K. Abe, R. C. Fernow, A. D. Krisch, T. A. Mulera, A. J. Salthouse, B. Sandler, K. M. Terwilliger, D. G. Crabb y P. H. Hansen en *Physical Review Letters,* vol. 39, n.º 12, págs. 733-736; 19 de septiembre de 1977.

LA CONQUISTA DE EUROPA OCCIDENTAL POR LAS LENGUAS INDOEUROPEAS

INDO-EUROPEAN AND INDO-EUROPEANS. PAPERS PRESENTED TO THE THIRD INDO-EUROPEAN CONFERENCE AT THE UNIVERSITY OF PENNSYLVANIA, dirigido por G. Cardona, H. M. Hoenigswald y A. Senn, Philadelphia 1970.

SPRACHEN UND INSCHRIFTEN. A. Tovar. Amsterdam, 1973.

FLEXION UND WORTBILDUNG, dirigido por H. Rix, Akten der 5. Fachtagung der Indogermanischen Gesellschaft, Wiesbaden, 1975.

KRAHES ALTEUROPÄISCHE HYDRONYMIK UND DIE WESTINDOGERMANISCHEN SPRACHEN. A. Tovar en *Sitzungsberichte der Heidelberger Akademie der Wissenschaften, Philos.–hist. Klasse,* 1977.

INDOGERMANISCH UND KELTISCH, KOLLOQUIUM DER INDOGERMANISCHEN GESELLSCHAFT IN BONN, dirigido por K. H. Schmidt, Wiesbaden, 1977.

LA BASE MOLECULAR DEL MOVIMIENTO CELULAR

CELL MOTILITY. Dirigido por R. Goldman, T. Pollard y J. Rosenbaum. Cold Spring Harbor Conferences on Cell Proliferation, 1976.

ACTIN, α-ACTININ, AND TROMPOMYOSIN INTERACTION IN THE STRUCTURAL ORGANIZATION OF ACTIN FILAMENTS IN NONMUSCLE CELLS. Elias Lazarides en *The Journal of Cell Biology,* vol. 68, n.º 2, págs. 202-219; 1976.

FLUORESCENT ANTIBODY LOCALIZATION OF MYOSIN IN THE CITOPLASM, CLEAVAGE FURROW, AND MITOTIC SPINDLE OF HUMAN CELLS. Keigi Fujiwara y Thomas D. Pollard en *The Journal of Cell Biology,* vol. 71, n.º 3, págs. 848-875; diciembre, 1976.

UTILIZACION DEL LASER EN QUIMICA

INFRARED LASER CATALYZED CHEMICAL REACTIONS. A. M. Ronn en *Spectroscopy Letters,* vol. 8, n.º 5, págs. 303-328; 1975.

CHEMICAL AND BIOCHEMICAL APPLICATIONS OF LASERS: vol. 3. Dirigido por C. B. Moore. Academic Press, 1977.

SELECTIVE DISSOCIATION OF POLYATOMIC MOLECULES BY INTENSE INFRARED LASER FIELDS. Rafael V. Ambartzumian y Vladilen S. Letokhov en *Accounts of Chemical Research,* vol. 10, n.º 2, págs. 61-67; febrero, 1977.

LASER AND CHEMISTRY. Sol Kimel y Shammai Speiser en *Chemical Reviews,* vol. 77, n.º 4, págs. 437-472; agosto, 1977.

INFRARED INDUCED UNIMOLECULAR REACTIONS. N. Bloembergen y E. Yablonovitch en *Physics Today,* vol. 31, n.º 5, págs. 23-30; mayo, 1978.

UN SISTEMA DE ENFRIAMIENTO DEL CEREBRO EN LOS MAMIFEROS

INFLUENCE OF THE CAROTID RETE ON BRAIN TEMPERATURE IN CATS EXPOSED TO HOT ENVIRONMENTS. M. A. Baker en *Journal of Physiology,* vol. 220, n.º 3, págs. 711-728; febrero, 1972.

RAPID BRAIN COOLING IN EXERCISING DOGS. M. A. Baker y L. W. Chapman en *Science,* vol. 195, n.º 4280, págs. 781-783; 25 de febrero de 1977.

PROBLEMAS INTRINSECAMENTE DIFICILES

THE DESIGN AND ANALYSIS OF COMPUTER ALGORITHMS. Alfred V. Aho, John E. Hopcroft y Jeffrey D. Ullman. Addison-Wesley Publishing Co., 1974.

WEAK MONADIC SECOND ORDER THEORY OF SUCCESSOR IS NOT ELEMENTARY-RECURSIVE. Albert R. Meyer en *Lecture Notes in Mathematics: No. 453,* dirigido por A. Dold y B. Eckmann. Springer-Verlag, 1975.

AN INTRODUCTION TO THE GENERAL THEORY OF ALGORITHMS. Michael Machtey y Paul Young. Elsevier North-Holland Publishing Co., 1978.

EL COMETA DE HALLEY PINTADO POR GIOTTO

THE NATURE OF COMETS. Nikolaus B. Richter. Methuen & Co Ltd., 1963.

THE COMPLETE PAINTINGS OF GIOTTO. Edi Baccheschi. Weidenfeld & Nicolson, 1969.

GIOTTO: THE ARENA CHAPEL FRESCOES. Dirigido por James H. Stubblebine. W. W. Norton & Co., Inc., 1969.

JUEGOS MATEMATICOS

MAGIC AND PARAPHYSICS. Martin Gardner en *Technology Review,* vol. 78, n.º 7, págs. 42-51; junio, 1976.

GELLER, GULLS, AND NITINOL. Martin Gardner en *The Humanist,* vol. 37, n.º 3, págs. 25-32; mayo/junio; 1977.

TALLER Y LABORATORIO

A COMPENDIUM OF SPHERICAL ASTRONOMY. Simon Newcomb. Dover Publications, Inc., 1960.

DOUBLING YOUR SUNSETS, OR HOW ANYONE CAN MEASURE THE EARTH'S SIZE WITH A WRISTWATCH AND METERSTICK. Dennis Rawlins en *American Journal of Physics,* vol. 47, n.º 2, págs. 126-128; febrero, 1979.